Interdisciplinary Contributions to Archaeology

Edward B. Banning

The Archaeologist's Laboratory

The Analysis of Archaeological Evidence

Second Edition

MOREMEDIA

Springer

Interdisciplinary Contributions to Archaeology

Series Editor

Jelmer Eerkens
University of California, Davis
Davis, CA, USA

More information about this series at http://www.springer.com/series/6090

Edward B. Banning

The Archaeologist's Laboratory

The Analysis of Archaeological Evidence

Second Edition

Springer

Edward B. Banning
Department of Anthropology
University of Toronto
Toronto, ON, Canada

ISSN 1568-2722
Interdisciplinary Contributions to Archaeology
ISBN 978-3-030-47990-9 ISBN 978-3-030-47992-3 (eBook)
https://doi.org/10.1007/978-3-030-47992-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

For Lyn and Jordan

Preface

A new edition of *The Archaeologist's Laboratory* has been long overdue. Not only have things changed in our discipline over the last two decades, I have long wanted to give the book better focus and to illustrate its concepts with recent case studies. Although it is impossible to separate lab work completely from fieldwork, the emphasis here is on the former, the activities that archaeologists and their colleagues carry out after at least some fieldwork is complete. However, we should always keep in mind that fieldwork and laboratory work go hand-in-hand, and the best research designs are not compartmentalized in one or the other.

A thread woven throughout the book is the quality and validity of archaeological arguments and the data we use to support them. No matter what our theoretical orientation, we want to make academic arguments and heritage recommendations that are convincing. At a bare minimum, we want to convince ourselves and others that the results we seem to have are not just due to measurement errors or poor research design. We want to draw inferences that are at least reasonable given the evidence available to us, and to present conclusions or hypotheses that arguably fit the evidence better than some competing hypotheses, and especially the competing hypothesis that apparent patterns are only due to sampling problems, investigator bias, or measurement error.

Consequently, I have aimed in this book to present basic information about many of the most common laboratory activities that archaeologists carry out, and a few less common ones, but in the context of ensuring, to the extent that it is possible, that these activities result in observations, measurements, and conclusions that are accurate, valid, and relevant to the research problems at hand. While it has been necessary to subdivide laboratory research into a number of arguably arbitrary categories, it is important to realize that archaeology is a holistic discipline that works best when we consider different kinds of evidence, such as pottery and plant and animal remains, in tandem. It would be impossible to understand past food and culinary practices properly, for example, by examining only one of these. That is just one of the reasons that archaeologists typically work in research teams that combine a variety of expertise.

I also firmly believe that the way we present our research results to others is just as important as the way we conduct research, so I devote space to the presentation of data, mainly in graphs and artifact illustrations, but with at least brief reference to 3D imaging and the media of archaeological dissemination, which today go well beyond traditional monographs and academic journals.

I have tried to give the book an international and inclusive flavor by including examples and case studies from around the world and from the Paleolithic to historical archaeology. Archaeology in different parts of the world has different challenges and different research problems that sometimes call for very different research strategies, even though many aspects of research are more universal.

This book has benefited from advice and suggestions I received from and conversations I have had with a great many students and colleagues. I owe some of my thinking on topics covered in this book to the early influences of Ron Farquhar, Ted Litherland, Ron Hancock, and the late Jack Holladay. In addition to those whom I thanked in my first edition, I would like to acknowledge valuable suggestions for improvements to one or more chapters by Caitlin Buck, Ben Collins, Costis Dallas, Tom Dye, Kevin Gibbs, the late Stuart Laidlaw, Lee Lyman, Danielle MacDonald, Lisa Maher, Joy McCorriston, Katherine Patton, Trevor Ringrose, Branden Rizzuto, Aaron Shugar, Jim Skibo, James Stemp, Susan Stock, Todd Surovell, John Triggs, and John Whittaker. I also would like to thank Teresa Krauss, Christi Lue, and Sowmya Thodur at Springer; Sophia Arts for her assistance with copy-editing and the index; and Emerson Grossmith, Seiji Kadowaki, Steve Rhodes, Joy McCorriston, Emma Yasui, and Emily Hubbard for photos of potting, core refitting, bone awls, charred rachis segments, starch and phytoliths. Arno Glasser, Steve Rhodes, and Stuart Laidlaw helped with some of the other photos, and William Wadsworth and Yara Salama drew many of the illustrations, while Julia Pfaff and Catharine Solomon provided several lithic illustrations. The remaining photographs and illustrations, unless noted otherwise, are by the author.

Toronto, ON, Canada Edward B. Banning

About the Book

The book has two main parts.

Part I introduces general topics concerning data, measurement error, data quality, research design, typology, probability, and databases. This section also covers presentation in graphs and tables and basic artifact conservation. Part I rounds out with the very important topic of laboratory safety.

Part II offers brief surveys of the analysis of various classes of artifacts, namely lithics and ground stone, pottery, metal artifacts, bone and shell artifacts, and what are sometimes called "ecofacts"—animal and plant remains and sediments—and dating by stratigraphy, seriation, and chronometric methods. The emphasis in the chronometric chapter is not on how dating methods work, but on how to interpret dating evidence, with focus on dendrochronology and radiocarbon. A chapter on archaeological illustration and publication completes this section of the book.

A book like this cannot cover everything, so you will not find in-depth treatment of any of these topics, while some kinds of laboratory research are necessarily absent. To give a taste of greater depth, I include some case studies to illustrate aspects of archaeological laboratory research or to emphasize one of the book's main themes, and readers will find some important or useful sources in the bibliographies at the end of each chapter, including both classic works and recent literature.

Throughout the book, words in bold indicate important terms and concepts, with bolding usually occurring where the concept is first defined. Algebraic terms for statistics and mathematical expressions are indicated in italics, with Greek letters for population parameters (e.g., σ) and italicized Roman letters (e.g., s) for samples. Estimates of parameters are indicated with a circumflex accent (e.g., $\widehat{\sigma}$).

I should also emphasize that this is not a statistics text, even though there are references to statistics in several chapters, Chap. 6 has an extended discussion of sampling, and Chap. 8 does review basic concepts regarding probability and statistics. I also introduce some statistical topics, like Bayesian analysis, that will be unfamiliar to some readers. As probability and sampling are fundamental aspects of modern archaeology, whether formally part of research designs or not, Chaps. 6 and 8 provide necessary background for many of the other chapters. However, these chapters most certainly do not replace books on statistics or archaeological mathematics, so readers should consult some of the more general statistical resources cited there and texts specifically aimed at archaeologists (e.g., Barcelo and Bogdanovic, eds. 2015; Baxter 2003; Buck et al. 1996; Drennan 2010; Shennan 1997). For those who find statistics particularly terrifying, I highly recommend Fletcher and Lock's (1991) *Digging Numbers* and Rowntree's (2018) *Statistics Without Tears*. However, students will derive greater benefit from this book if they already have at least some basic background in statistics, as this book only presents a brief review before moving on to more complex topics.

Some chapters also include a little mathematics, which some students may find off-putting. My reasons for including this are to make the point that the equations are not mysterious and to increase students' comfort with at least simple algebraic and statistical or probability expressions. I try to follow each such equation with a verbal explanation of what it means and, in some cases, an example to illustrate how it works.

Although some of the topics covered, especially in Part I, can involve statistical or database software, I have generally avoided discussing particular software platforms because those change so rapidly and I do not want to give the impression that I am endorsing any particular product. However, some instructors may want to introduce students to one or more of the widely available software products, such as R or MySQL, in conjunction with some of the topics covered here. The exception to my rule is that, in Chaps. 19 and 20, I do mention several of the commonly available software options for ordering contexts stratigraphically and for interpreting radiocarbon data, many of which are available to users at no cost.

The text varies in level of difficulty. Generally, most chapters begin with fairly basic concepts and descriptions of terminology that would be appropriate for undergraduate students, perhaps in second or third year, who have had at least some background in archaeology. Some later parts of chapters explore concepts that might be a little advanced for such students, sometimes because they are more quantitative, sometimes just because the topics are more specialized. The last sections of some chapters briefly summarize a few major research questions that are relevant to the chapter, along with some case studies. The variation in level of difficulty allows the book to be used for senior undergraduates and graduate students for whom the more basic parts of each chapter may constitute review, but who would benefit from the greater level of detail or more challenging concepts in later sections. Such students would also benefit from reading sources cited there.

I encourage instructors who use this book for course readings to select portions of chapters that are most appropriate for their students' level of expertise. I would also not expect students necessarily to read the book in its linear order. I have found it effective to alternate between the topics in Parts I and II so that students can learn the more general or theoretical material in the context of particular kinds of materials or research problems. For example, it may be useful to cover quantification in the context of zooarchaeology if you have access to an osteological teaching collection.

When I teach the topics covered in this book, I have students carry out laboratory activities, such as measuring artifacts and comparing results to demonstrate inter-observer variations, or cluster sampling modern seeds in bins full of sand. Some of these lead to formal labs that students must submit for grading. I have posted a few examples of these laboratory assignments online at http://extras.springer.com/

It is also my hope that students will benefit from returning to the text later in their careers to refresh their memories of important concepts or for direction to more specialized literature relevant to their current research, whether that is in academia, government, or the private sector. Graduate students and professional archaeologists may also find this book a useful reference, with the index guiding them to relevant pages.

Toronto, ON, Canada Edward B. Banning

References

Barcelo, J. A., & Bogdanovic, I., Eds. (2015). *Mathematics and archaeology*. Boca Raton: CRC Press.

Baxter, M. J. (2003). *Statistics in archaeology*. New York: Wiley.

Buck, C. E., Cavanagh, W. G., Litton, C. D. (1996). *Bayesian approach to interpreting archaeological data*. New York: Wiley.

Drennan, R. D. (2010). *Statistics for archaeologists. A common sense approach*. New York: Springer.

Fletcher, M., & Lock, G. R. (1991). *Digging numbers: Elementary statistics for archaeologists*. Oxford: Oxford University Committee for Archaeology.

Rowntree, D. (2018). *Statistics without tears: A primer for non-mathematicians*. London: Penguin Books.

Shennan, S. (1997). *Quantifying archaeology* (2nd ed.). Edinburgh: Edinburgh University Press.

Contents

About the Author

Ted Banning is a Canadian archaeologist and professor of Anthropology at University of Toronto. His research focuses on the Neolithic of the Near East, but his other interests include the theory and methods of archaeological survey and landscape archaeology, the spatial organization of ancient built environments, and the spatial and statistical analysis of microrefuse —tiny fragments of lithics, pottery, ground stone, shell, bone, and other materials in archaeological sediments. For more than three decades, he has directed field projects in northern Jordan, and previously served on the staff of excavations at the Pre-Pottery Neolithic site of ʻAin Ghazal, Jordan, at Tell al-Maskhuta, Egypt, and on surveys in southern Jordan and northeastern Egypt. He also has side interests in historical archaeology and numismatics, and especially the early history of numismatists and antiquarians in Canada. He is author of *Archaeological Survey* (Springer 2002), a member of the Society of American Archaeology and American Institute of Archaeology, and Fellow of the Society of Antiquaries of London.

List of Figures

List of Tables

Part I

Introduction to Part I

Keywords
Archaeological interpretation; Quality Assurance; Data quality assessment

One of the basic purposes of this volume is to introduce students to terminology and some of the methods that archaeologists use to analyze artifacts and other remains that they have brought back from fieldwork.

A main theme throughout the volume is that all archaeological interpretation, no matter its theoretical orientation, depends on observations on artifacts and other things that we can describe, in some sense, as measurements. As Chap. 1 and later chapters develop more fully, it is impossible for us to make any such observations entirely objectively, or without error. Sometimes the error results from the imprecision of our measuring instruments, such as balances and calipers. Sometimes it results from differences between people; archaeologists and lab assistants looking at the same thing see it in different ways. And sometimes it is due to less obvious but still important factors, such as preservation in the ground, variations intrinsic to the materials, statistical error, or even failure to account for impacts of artifact collectors. If we expect to base confident conclusions on our observations, we need to understand these influences on our measurements. More fundamentally, even our choices about what kinds of observations to make and how to make them have a substantial impact on our "results," and these choices depend in large part on our theoretical perspectives.

The chapters of this book fall into two parts. Part I covers topics that are applicable to most or all kinds of archaeological laboratory work, such as measurement, error, research design, systematics, and quantification. There are also brief summaries of the important topics of conservation, lab management, and lab safety. Part II instead deals with different categories of materials and chronological data that archaeologists analyze, as well as with illustration and publication.

Quality in Archaeological Data

A theme that runs throughout these chapters is the quality of the data and the inferences we base on data. If all archaeological data have errors and uncertainties, and are also "theory-laden," how can we be confident of any conclusions we draw from archaeological observations? Archaeologists have thought about this problem for a long time (Daniels 1972), but it still receives less attention than it should. It is far from unique to archaeology, however, and affects so-called empirical and so-called interpretive archaeology equally.

In many fields, a way to manage these uncertainties is Quality Assurance. Quality Assurance (QA) consists of policies, procedures, manuals, standards, and systems that help us ensure and improve the quality of products or services and their users' satisfaction with them. Most of the literature on QA focuses on its application in manufacturing or service contexts, but it is also applicable to the quality of archaeological measurements, analyses, and interpretations. QA is one of the threads woven through this volume, counterbalancing the focus on measurement error.

Any well-run archaeological laboratory will make use of policies or protocols that guide the way its students or employees are to carry out typical procedures, such as washing, labelling, conserving, classifying, counting or measuring artifacts and other materials of interest to archaeologists. It will also have protocols for illustrating things, for backing up computer files, for curation of digital data, for lab health and safety, for ensuring compliance with laws and government guidelines, and probably a host of other things. Students rarely see explicit reference to such policies in textbooks, but they are essential to the quality of the observations on and inferences about those materials, as well as the health and safety of people in the lab.

In some labs, the main Quality tool is a manual that describes the procedures for such things as labelling washed artifacts and ensuring that their archaeological context is accurately recorded. Most such manuals also describe the lab's preferred way to measure such things as artifact size, and some include pictures of typical artifacts belonging to each category of a classification system. Most archaeologists are very familiar with the preparation and use of such manuals, even if they do not think of it as Quality Assurance.

In heritage management or Cultural Resource Management (CRM), quality work also involves protocols to ensure compliance with relevant laws and government standards and guidelines. Many firms that do archaeological heritage assessments employ templates to make sure that their technical reports cover all of the required topics and in the appropriate formats. For such firms, QA also needs to cover compliance with employment standards, tax law and other business-related things that an academic lab might not have to consider.

Today, there is an increasing emphasis on QA in archaeological work both in the field and in the lab (e.g., Banning et al. 2017; Baylis 2015; Hunt and Sadr 2014; Nims and Butler 2017; Wolverton 2013). It is important for students to learn some of the tools for evaluating the quality of research already completed. This aspect is known as quality assessment. Another aspect is quality control, which has the goal of preventing or minimizing certain kinds of errors in future.

Throughout this book, and especially in Chap. 1, you will find reference to these and some more complex tools that can help us ensure the reliability of our archaeological observations and the inferences we base on them.

Summary

- No matter your theoretical orientation, archaeological data have errors and are subject to interpretation
- It is important to manage these errors and other sources of variation in analysis in order to draw plausible inferences from data, or to evaluate competing hypotheses
- Quality Assurance (QA) is a suite of tools to help ensure the quality of any "product," including data and interpretations of data
- Quality assessment is that aspect of QA that, in an archaeological context, involves evaluating work to see if it meets standards or expectations, or if the data include biases that could negatively affect our results
- Quality control is another aspect that involves tools that help us prevent or minimize errors in future. This can include things as simple as manuals and guidebooks, or reference collections, but can also include periodic audits or spot-checks by experts

References Cited

Banning, E. B., Hawkins, A. L., & Stewart, S. T. (2017). Quality Assurance in archaeological survey. *Journal of Archaeological Method and Theory, 24*(2), 466–488.

Baylis, A. (2015). Quality in Bayesian chronological models in archaeology. *World Archaeology, 47*(4), 677–700.

Daniels, S. G. H. (1972). Research design models. In D. Clarke (Ed.), *Models in archaeology* (pp. 201–229). London: Methuen.

Hunt, T., & Sadr, K. (2014). Inter-analyst variability in identification and classification of pre-colonial stone-walled structures using Google Earth, southern Gauteng, South Africa. *The South African Archaeological Bulletin, 69*(199), 87–95.

Nims, R., & Butler, V. L. (2017). Assessing reproducibility in faunal analysis using blind tests: A case study from northwestern North America. *Journal of Archaeological Science: Reports, 11*, 750–761.

Wolverton, S. (2013). Data quality in zooarchaeological faunal identification. *Journal of Archaeological Method and Theory, 20*(3), 381–396.

1 What Are Data? Measurements and Errors

I not only have been unable to use other investigators' data, but I have also frequently found my own data lacking in many important 'facts'—facts which could have been collected had I been aware of the questions to which the given observation was relevant

(Binford 1964: 427).

For students of archaeology, it is tempting to assume that archaeological data are obvious, or just "are." It would be more accurate to say that they are what we make them, but that's not the whole story either. Observation is the archaeologist's "basic tool," but there is more to it than just "looking" (Kehoe and Nelson 1990). This chapter—and this book more generally—will develop concepts that archaeologists, consciously or not, employ while observing, recording, and analyzing things. There is not just one way to make an archaeological observation or measurement, or to present it to others, and no measurement we can make is ever immune to error. Good archaeology is all about making choices, including our choice of the right measure to study some phenomenon of interest, how to display those measures, and how much confidence to place in them.

This book uses "measurement" in a very broad sense to include any kind of observation an archaeologist might make. The description of color, for example, is a measurement that archaeologists often measure on a standardized scale. Measurement theory provides a way of describing things and making them comparable to one another. Other tools can help us summarize, display, and share the data we have collected or measured. But these tools do not guarantee objectivity. Even though professional archaeologists are skilled observers, their observations depend on guiding assumptions and preconceptions (Kehoe and Nelson 1990: 1).

From the initial collection of data through their analysis and publication, there is always at the very least a selection of data from a theoretically infinite number of possible observations (Coombs 1964). Although some archaeologists might tell you that you should try to collect data "objectively," in any science, as in daily life, we are only able to observe something if we are prepared to see it.

Paraphrasing Henri Poincaré, the novel, *Zen and the Art of Motorcycle Maintenance* makes this point:

> According to the doctrine of 'objectivity,' . . . . We should keep our mind a blank tablet which nature fills for us, and then reason disinterestedly from the facts we observe.
>
> But when we stop and think about it disinterestedly, ... Where are those facts? What are we going to observe disinterestedly? . . . The right facts, the ones we really need, are not only passive, they are damned elusive, and we're not going to just sit back and "observe" them. We're going to have to be in there looking for them. . . . (Pirsig 1974: 274–75, italics in original)

Data are not objects, such as projectile points or potsherds. They are observations and measurements we decide to make on things like points and potsherds, and on their archaeological contexts—in what sites or layers we find them, with what other artifacts they co-occur, under what environmental circumstances they accumulated. We do not even just select facts from an infinite "sea" of data but *construct* them by deciding *how* we will "see" them. Data are theory-laden and their observation is historically contingent; 100 years ago, no archaeologist would have considered the abundance of ^{14}C in a piece of charcoal to constitute meaningful data. That does not mean that things do not exist unless someone observes them; but anything we are not prepared to observe cannot serve as data.

Archaeological data can include such things as the name or category of an artifact (e.g., "Clovis point"), the density of potsherds on the surface of a site (12 sherds per square

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47992-3_1) contains supplementary material, which is available to authorized users.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_1

meter), the average distance between a site and its nearest neighboring site in kilometers, the invasiveness of retouch on a stone tool, the order in which a potter incised design elements on a pot, or the "date" of a piece of charcoal in radiocarbon years BP. It can also include the spatial provenience of an artifact either by its *xyz* coordinates in a mapping system or by its stratigraphic context (e.g., layer 6), or practical information and details of method, such as the mesh size of screens used in excavation, the volume of sediment excavated, the spacing of shovel tests, or the thoroughness of field survey by a crew of fieldwalkers. It can include geological information and data on what archaeologists call site-formation processes. In each case, the type of measurement we use depends on the kinds of questions archaeologists are asking, such as "what is the age of this site?" or "what is the social meaning of this decoration?"

The trick is deciding how to measure these things. One might argue that our perceptive abilities, interests, social, economic and cultural backgrounds, and even our unique historical contexts, so pervade our choice of questions and ways of obtaining answers that we cannot ever hope to draw reasonable conclusions about people in the past. I think this is too pessimistic. If nothing else, each archaeologist's interpretation makes a sort of target that other archaeologists will re-evaluate and sometimes reject, sometimes by collecting and analyzing kinds of data that the first archaeologist didn't think were important. This process might, in an iterative fashion, lead to better and better interpretations.

But other archaeologists would argue that we don't need to expect just one interpretation to be the "right" one. The concept of multivocality embraces the idea that different people—both now and in the past—look at and express the same thing in different ways.

No matter whether we belong to a school of thought that some theories and interpretations are measurably better than others, or to one that embraces multivocality, as archaeologists we do expect one another's theories to be "grounded" in data, and we have some reasonable expectations about the quality of the data. To use an extreme example, the data should not just be "made up," but rather be based on careful and thoughtful observations of relevant archaeological phenomena, even though we recognize that those observations are theory-laden.

Some authors (e.g., Shanks 1992: 26) have suggested an opposition between a "scientific" archaeology that gives primacy to facts and measurements, which are supposed to be neutral or objective, and other archaeologies that emphasize interpretation. This is a false dichotomy. Even in "hard" sciences, such as physics, there is widespread recognition that data are theory-laden and the kinds of observations that scientists make depend on the instruments available and the kinds of theory currently on the forefront. Einstein struggled with the failure of deduction and induction to account for the advances in science (de Waal 2016). At the same time, practitioners of interpretive or postmodernist archaeologies try to ground theories in data as diverse as structure in the ordering and orientation of decoration (Shanks and Tilley 1987: 160), the petrography of pottery (Jones 2002: 120–122), the spatial organization of buildings (Thomas 2013), and the topography and "viewsheds" in a landscape of monuments (Tilley 1994).

Our research questions and other motivations guide what kinds of measurements we think are worth measuring. Sometimes we select certain kinds of measurements because we think they will help us answer our own research questions; sometimes we measure things because it helps us collaborate with others' projects. Sometimes certain measures are required to satisfy regulatory requirements, especially in Cultural Resource Management and private-sector archaeology. And sometimes we may record certain kinds of information that we think will help future archaeologists understand our fieldwork or laboratory analyses.

Archaeologists of all stripes make use of data and measurement, and measurement theory guides us in their collection and characterization. All data are filtered through the observer's senses, instruments, and theoretical predispositions. We have to measure things with our eyes, measuring tapes, calipers, artifact typologies, conceptual categories, and other instruments and conceptual tools ranging from crude to sophisticated. No system of observation can be theory-neutral (Hanson 1958).

Measurement consists of comparison. You make an observation by comparing some object with a scale, whether that scale is a ruler, measuring tape, Munsell chart (p. 298), artifact typology, or cultural categories. Measurement consists of assigning an abstract symbol—a number and unit, a color, a name, an icon—to represent the object or the value or magnitude of one of its attributes or qualities. Conventionally, archaeologists, like other social and natural scientists, do this by reference to one of several scales of measurement.

1.1 Scales of Measurement

Archaeologists commonly distinguish **qualitative** from **quantitative** research. Qualitative measurements consist of assigning observations to categories. Quantitative measurements instead represent magnitudes. As the qualitative-quantitative distinction is really too crude to characterize all the ways that we can make measurements, however, it is common to use more nuanced categories to describe measurements themselves. The conventional names for these categories are nominal, ordinal, interval and ratio scales of measurement.

A **nominal scale** consists only of unordered categories, such as kinds of pottery decoration or vessel shapes, and these categories are "unweighted," which means that no category is more important than any other. The simplest kind of nominal scale (a "subcategory" if you will) is the **dichotomous** scale, which is a nominal scale with only two categories, such as male and female, large and small, or present and absent. Nominal scales are very important in archaeology because archaeologists make so much use of classification and typology (Chap. 3), which are all about nominal scales.

Nominal scales are also essential if you want to count anything; counting by its nature implies that you are assigning items to categories. The process of counting how many belong to each category is called **enumeration**. For example, we might count how many sherds are decorated and how many are plain, or how many charred seeds belong to maize, chenopods, fleshy fruit, and so on. Enumeration is the first step for several other kinds of measures, and particularly for measuring proportions, percentages and densities.

An **ordinal scale** also consists of categories but, unlike those of nominal scales, these ones are ordered, as in "small, medium, large," "early, later, recent," or "rare, common, ubiquitous." Mathematically speaking, ordinal scales allow deductions of the sort: if A is greater than B and B is greater than C then A is greater than C. Archaeologists make considerable use of ordinal scales to categorize artifacts or sites by size, but by far the greatest use of ordinal scales is in chronology. Archaeological time periods, such as "Archaic," "Middle Woodland," "Lower Palaeolithic," "Neolithic" or "Early Bronze Age," are all categories on ordinal scales. Stratigraphic information is also related to ordinal scales; it is the sequence or order of layers and other events that matters (see Chap. 19).

A **ranked scale** is a special type of ordinal scale in which each category includes only one member (except where there are exact ties). For example, we can rank a bunch of projectile points from smallest to largest on the basis of their greatest dimension or their axial length (see p. 164).

In addition, some fields make use of a kind of scale that, in this classification, would be ordinal but whose users treat almost like an interval scale. Likert (1932) introduced these scales to psychology, with five or seven options ranging from "strongly agree" to "strongly disagree" or "almost always true" to "almost never true" to measure people's attitudes. **Likert-like scales** are often used in opinion surveys and health assessments, but only rarely in archaeology (e.g., Proulx 2013: 115; Sifniotis et al. 2007). Conceivably, they could have a role in fuzzy classification (p. 37) and estimating prior probabilities for Bayesian analysis (pp. 136, 139). While it is not uncommon for users of such scales to summarize data with means, it is more appropriate to use the median as a measure of their central tendencies in data like this, as these are not really interval scales.

Nominal and ordinal scales are qualitative in the sense that, even if we label their categories with numbers, those numbers are just names, not magnitudes. On an ordinal scale, we can say that "medium" sites are larger than "small" sites but cannot say by how much. Nor can we assume that the projectile ranked sixth is twice as large as the one ranked third. Most likely, that would be wildly incorrect.

When people talk about quantitative measurements, they usually mean interval scales and ratio scales.

An **interval scale** not only has inherent order, like an ordinal scale, it also has constant intervals or distances between the points on that scale, and typically these intervals have named units. That means the difference between, say, 5 and 7 on the scale is the same as the difference between 7 and 9 or 100 and 102. This property makes it possible for us to do the arithmetical processes of addition and subtraction.

However, interval scales do not necessarily allow multiplication and division because they don't always have a "real" zero point. For example, the zero on the most common scales of temperature is purely arbitrary, and neither 0° Celsius nor 0° Fahrenheit means an absence of temperature. They're just arbitrary benchmarks. Consequently, it makes no sense to say that 30° is twice as hot as 15° in either of those systems.

To permit multiplication and division, you need a special kind of scale that not only has constant intervals, but also a real zero point that means a total absence of something. That is a **ratio scale**. A common example of a ratio scale is a scale of height or length or distance, as measured in centimeters or meters or miles or cubits. Anything that is 0 m long has an absence of length (i.e., doesn't exist physically at all) and anything that is 10 m long is twice as long as something 5 m long. Again, ratio scales often have standard units, such as meters, square meters, grams, hours or minutes, or degrees of angle. However, when we count things or measure densities, we also use a ratio scale. The things we're counting are classified on a nominal or ordinal scale, but the actual counts are ratio-scale data. Densities are just counts per some other unit, for example the number of artifacts per square meter or per cubic meter, or number of droughts per century.

There is one other aspect of interval and ratio scales that we need to consider. All of them have constant intervals but, in some cases, observations are discrete and in others they are continuous. When we count things, for example, we have **discrete measurement**, meaning that our intervals are always whole numbers. Your family can have 3, or 7 or 10 members, but never 2.7 members. Of course, you've often seen statistics such as "mean family size is 2.7 people" but that doesn't mean there is any actual family with 2.7 people. Counting people is a discrete measurement. Likewise, a stone tool could have scars from one, two, three, or 15 previous flake removals, but not from 7.8 flake removals, even if the average number of flake removals for a whole collection of stone tools might be 7.8.

Nominal and ordinal measurements are also discrete, since you do not expect any "grey areas" between categories (although see "fuzzy classification," p. 37).

When we measure the length of a projectile point with ruler or calipers (Fig. 1.1), by contrast, we have continuous measurement. Whatever units we use—inches, millimeters, or even miles—we can theoretically measure to as many

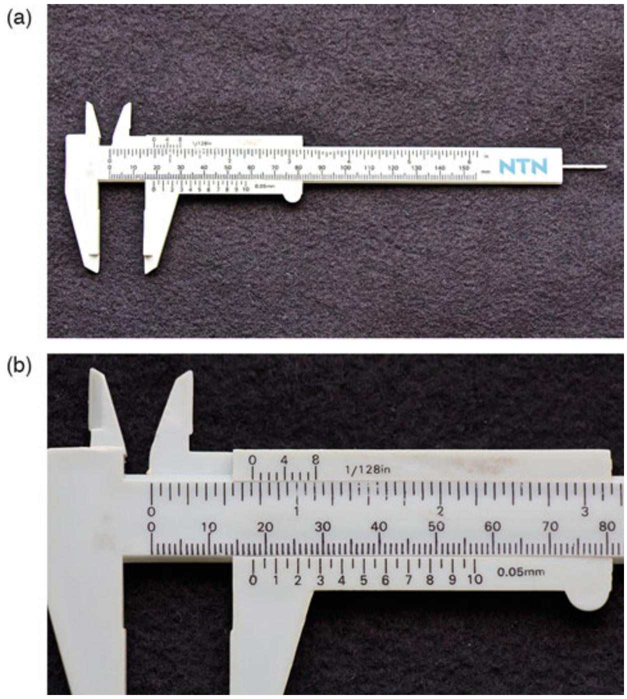


Fig. 1.1 Manual caliper and close-up of the measuring scale. Note how the "zero" tick on the mm scale falls just a little to the left of 18 mm. That suggests that the measurement should be close to 17.7 or 17.8 mm. Now, looking along the bottom scale (the 0.05 mm scale), the tick that best lines up with the tick above it is the one between "7" and "8," indicating that the measurement is 17.75 mm (0.05 is its resolution or precision). Given that this is a rather cheap and possibly not highly accurate caliper, one might be more comfortable citing 0.1 mm resolution (or 17.8 mm)

decimal places as we want. Thus, the projectile point might be 46.7 mm long, or 1.84 inches, or 0.000029 miles, or 46.6891387 mm.

The defining characteristic of a continuous scale is that, if you take any two points along the scale, you can *always* find other measurements between them. For example, between 46.7 and 46.8 mm you can find 46.74. Between 46.74 and 46.75 you can find 46.749. Between 46.749 and 46.75 you can find 46.7497, and so on.

This is a very important characteristic, as it means that no measurement we make on a continuous scale is ever exact. No matter how carefully we make a measurement on an artifact, someone else could always come along, re-measure it and arrive at a slightly different result, especially if they used either a less precise or a more precise measuring instrument. This brings up the concepts of accuracy and precision in measurements.

1.2 Measurement Errors and Uncertainty

As just seen, the very nature of continuous measurement means that we can never measure things exactly on a continuous ratio or interval scale. The value that results from our measurement depends on a number of things, including the kind and quality of our measuring instrument and our own decisions about how "careful" our measurement needs to be. Some measurement errors are small and possibly random, others are large, surprising, and perhaps a little embarrassing, while others are systematic so that they could eventually lead to erroneous conclusions (Daniels 1972: 204–224; Heilen and Altschul 2013).

1.2.1 Precision

The simplest aspect of measurement error is the smallest increment that the measuring instrument is capable of distinguishing, which is technically its measurement resolution, but is popularly called **precision**. For example, some electronic balances will measure the mass of a projectile point to the nearest 0.1 g, and others to the nearest 0.001 g. If there is a good reason why we need to need to know the artifact's mass to the nearest 0.001 g, then we need to buy a more precise (and generally more expensive) balance. But there are many instances where we don't even need 0.1 g precision—for example, to "weigh" a whole bag full of pottery sherds—in which case we might content ourselves with an inexpensive balance that measures to the nearest 2 g. That is a decision we must base on how important we think a small difference in the mass of sherds would be (or how much error we think is acceptable). Would an error on the order of 5 g cause us to make a terrible mistake in the interpretation of an archaeological context? No textbook can answer that question for you; you need to make that decision on the basis of your own familiarity with the research problem at hand (see Chap. 6).

This variability in the resolution of measuring instruments of course affects the precision of the resulting measurements. For example, if we had to measure a projectile point's length with a 12-inch ruler with no intervals smaller than a quarter-inch, we might be able to say that it is 3-1/4 inches long, but it probably would not be reasonable to claim that it was 3-3/16 inches long. We might remeasure the point with cheap calipers, with a resulting measurement of 82.5 mm, and then measure it yet again with electronic calipers that tell us it is 82.477 mm long. Thus, the simplest expression of instrument precision is just the size of the smallest increment we can measure. In that event, we could express these several measurements as 3¼ ± ¼ in, 82.5 ± 0.1 mm, and 82.477 ± 0.001 mm. For 82.5, what the "±0.1" means is that we believe the "true" measurement lies somewhere between 82.4 and 82.6 mm.

Precision is the factor that should usually be most important in our decision about how many **significant digits** to show. Many students abdicate their responsibility for this to their calculators, smart phones or spreadsheets but, just because a device can show the result of some statistical analysis to ten decimal places doesn't mean that *you* should. When you make statements like "mean blade length is 51.327496 mm," you're suggesting that you actually measured the blades to ±0.000001 mm. Most likely, you did not. Furthermore, there is no plausible reason why you should, at least for most archaeological research problems. If, in fact, you only measured the blades with a precision of ±0.1 mm, then you should round off any calculations based on the measurements to reflect this. Consequently, the correct way to show your result would be 51.3 mm.

However, "significant digits" does not mean "decimal places." Decimal places depend too much on the units in which you express your measurements. For example, 2.1 km is a measurement with two significant digits, meaning that the distance it represents is somewhere between 2.0 and 2.2 km. 5.6 mm also has two significant digits and, like 2.1 km, has one decimal place. However, if we express these in different units, we have 2,100,000 mm and 0.0000056 km. One of these now has no decimal places, the other has seven decimal places, but *they both still have two significant digits*. Changing the units does not magically change the precision of the measurement.

The most straightforward definition of significant digits is that *they consist of all the certain digits plus the first uncertain one*. So, for 2.1 km, we are certain about the 2 (we know it's not 3 km or 1 km) but we can't be sure of the 1 (because measuring to three significant digits might result in 2.07, for example).

One way to be sure how many significant digits a number has is to convert it to scientific notation. In scientific notation, 2.1 km is $2.1 * 10^3$ m, while 5.6 mm is $5.6 * 10^{-3}$ m. Both of these have two significant digits (everything before the "*"). Meanwhile, $5.60 * 10^{-3}$ m has three significant digits because the trailing zero has no purpose other than to show us that the "true" measurement is somewhere between $5.59 * 10^{-3}$ m and $5.61 * 10^{-3}$ m. Table 1.1 provides other examples to show how this works.

The statistical definition of precision, however, has to do with how replicable the measurement is, that is, how much variability we see when we repeat the measurement many times. Generally, we attribute this variability to random error, but in practice it can come from many sources.

Table 1.1 Examples of measurements and their significant digits

Measurement	Scientific notation	Significant digits
26.01 cm	$2.601 * 10^1$	4
26,010 g	$2.601 * 10^4$	4
0.0026 kg	$2.6 * 10^{-3}$	2
14 artifacts/m^2	$1.4 * 10^1$	2
14 cm	$1.4 * 10^1$	2
14.00 cm	$1.400 * 10^1$	4
140.0 mm	$1.400 * 10^2$	4
101 °C	$1.01 * 10^2$	3
100 °C	$1 * 10^2$	1

1.2.2 Reliability

A concept closely related to precision in this sense is reliability. **Reliability**, or reproducibility, is the extent to which a measurement gives the same result in different situations, such as when different researchers do the measuring. Technically and formally, it is the proportion of the total variability in the measurement that is due to actual variability in what we are measuring. The remainder is variability due to such things as inter-observer differences and instrument problems. For example, when we need to make measurements on large numbers of lithics, we might divide the work among several student observers. But different people might vary in their carefulness, their eyesight, and their overall skill at using calipers or characterizing retouch. Consequently, you might expect some of the variation to be due, not to the actual characteristics of the lithics, but to the people who made the observations. There are ways to compensate for this effect in good research design (chap. 6, Daniels 1972: 215–225).

More formally, reliability has to do with the consistency or repeatability of a measurement. It is based on the "true score" theory, which models measurements as consisting of two main parts, the "true level" of the measurement and a component of (random) error. In highly reliable measurements, that error component should be small. Technical error of measurement (TEM) is a measure of replicability in repeated measurements. It is

$$R = \frac{s_T^2}{s_T^2 + s_e^2}$$

where s_T^2 is the sample variance of the mean of all the measurements (both sets), and s_e^2 is the variance on the mean differences between two sets of measurements (e.g., original and repeated), and R ranges up to 1.0 (perfect reliability). For an example of its use, see p. 261.

1.2.3 Bias

However, precision and reliability do not tell the whole story. Usually we want our measurements to be, not only reasonably precise, but accurate. **Accuracy** concerns the amount of **bias** in a measurement. Bias means systematically recording observations that (on an ordinal, interval or ratio scale) are systematically higher or lower than they should be. If, for example, you measured some stone tools with calipers whose edges were worn down or covered with tape, or measured the size of an archaeological feature with a cloth tape that had stretched, you would not get accurate measurements. You might consistently over- or underestimate the thicknesses of the stone tools by a constant amount (e.g., the thickness of the tape) or underestimate the width of the feature by some percentage that depends on the amount of stretching. A common source of instrumental bias is failure to **calibrate** electronic balances with standard weights at frequent intervals. The amount of this effect is called the bias, and an accurate measurement is reasonably free from bias, meaning that it is close to the true value.

Bias can also happen on nominal scales, however. For example, problems of preservation might make it easy to mistake the rim of a large jar as the rim of a bowl, but rarely the reverse. A result could be an over-estimate of the proportion of bowls.

Precise measurements need not be accurate, nor accurate ones precise. These are two quite separate aspects of the potential for error in a measurement (Fig. 1.2).

1.2.4 Errors of Classification

Up to now, we have focused mainly on errors in interval and ratio scales. Errors can also occur on nominal scales, a problem we call **misclassification**. In cases like that, it is more difficult to conceive of bias as some hypothetical difference between our measurement and a "true" value, and perhaps more sensible to think in terms of uncertainty or our degree of confidence in the observation. We can also consider the reliability of assignments to categories; how consistently do analysts assign a particular item to the same category? To what extent do the reported numbers of items in each category reflect the actual distribution, as opposed to classification errors?

The simplest nominal-scale assessment of error is the **error rate**, which is simply the proportion of classifications that are incorrect (see "joint probability of agreement," p. 14). However, that ignores the fact that poor measurers can make some correct classifications just by accident and implies that all kinds of misclassification are equally serious, which is generally not the case. For example, it may be less serious to misclassify a side-notched point as a corner-notched point than to misclassify it as a scraper. More complex ways of dealing with classification errors take these kinds of issues into account (see Hand 1997).

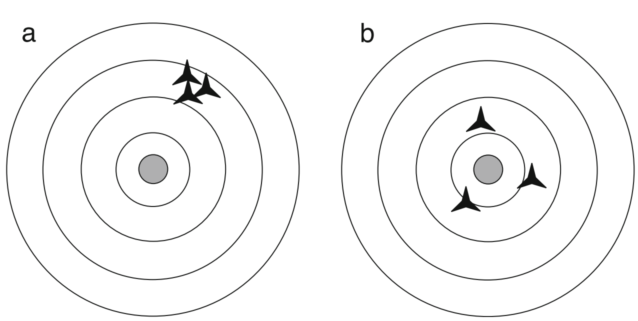


Fig. 1.2 The three triangles in (**a**) indicate three darts aimed very precisely but not very accurately, while the three in (**b**) are fairly accurate but less precise because they are not as tightly clustered

1.2.5 Cumulative Errors

Errors can also be **compensating errors** or **cumulative errors**. If, for example, you try to measure the area of a site by overlying a map of the site with a grid and counting squares, and count all the squares that are either within or on the site boundary, you will overestimate site area by counting parts of squares that lie outside the boundary. That is cumulative error. If instead you only count squares that seem to be more than half within the boundary and ignore the ones that more than half outside the boundary, your errors tend to cancel each other out, or compensate. The result of the former will have significant bias, while the latter will likely be fairly accurate, although not necessarily precise or reliable. Even greater bias would occur if you tried to estimate site area by simply multiplying the site's length by its width; almost always, that would grossly overestimate the site's size (large bias).

1.2.6 Outliers

Sometimes, a set of data contains one or two measurements that are very surprising, so different from the rest of the data that we find it hard to believe that they could be correct (Barnett and Lewis 1994). We call these **outliers**. Sometimes it is fairly obvious that they result from human error, such as making a mistake while copying data from field notes into a computer record, or a failure to read calipers correctly. When we can plausibly interpret outliers in this way, we can simply omit them from the data set or make new measurements to check on the surprising readings. One common practice has been to identify any observation more than three standard deviations from a sample mean as an outlier, but this is a problematic approach because the mean and standard deviation, themselves, are sensitive to outliers. A better approach is to use the median and median absolute deviation as the basis for identifying outliers (Ley et al. 2013).

In other cases, however, it is not so obvious that the outlier is just the result of human error. As you will discover in later chapters, even in well-behaved data we can expect extreme values to occur from time to time, just for statistical reasons.

It is important to note that precision and accuracy are terms associated, not only with measurements, but also with statistics, such as averages, a point to which we will return in Chap. 8.

1.2.7 Other Sources of Error

You should not assume that the measuring instrument, such as a ruler, tape, or caliper, is the only or even most important source of error in measurements. If it were, that simple conception of error as the smallest increment on the measuring device would be pretty reasonable. However, this leads to a false sense of accuracy and precision.

Suppose, for example, that you are involved in an archaeological survey and the survey crews measured the length and width of each site they discovered with steel tapes that allow you to measure quite easily to the nearest centimeter. Would it be appropriate or useful to claim that a site you measured was 93.16 m in length? Here, the main stumbling block to the accuracy and precision of your claim is not the measuring tape, it is your crew's ability to identify the edges of the site accurately and reliably. Most archaeological sites, unless they're walled cities, have very indistinct edges, such as a gradual diminishment in the density of scattered artifacts, or the slope of a mound, and no two archaeologists are likely to agree on the exact location of the site's boundaries or on the orientation of its long axis. Another archaeologist measuring the site you thought was 93.16 m long, even if she used exactly the same tape, might arrive at 106.05 m or 88.70 m. Even though the measuring instrument, the tape, is precise, the measurement is not very reliable. Consequently, it is not reasonable to suggest that the site was measured to four significant digits. In this instance, it would be more plausible to say the site was 95 or perhaps even 100 m long to be more honest to those reading your survey report.

1.2.8 Propagation of Errors

One of the things that happens when we carry out arithmetic operations on numbers that have errors associated with them is that the error on the number that results is a little bit bigger than the errors we start with. Fortunately, for most of the operations that archaeologists deal with, it is not very difficult to combine, or **propagate**, these errors (Peralta 2012).

If we add or subtract interval- or ratio-scale numbers that have errors, the way to propagate the errors is to square them, add the squared values together, and then take the square root of the result. For example, if we want to add two length measurements, 10.0 ± 0.5 mm and 5.3 ± 0.5 mm, the sum would be 15.3 mm and the total error would be the square root of $(0.5^2 + 0.5^2)$, or about 0.7.

For multiplication and division (including ratios), the process is only slightly more complicated. In these cases, we also sum the squared errors but, instead of using the absolute error values as in the example just given, we use relative errors. What that means is that we divide the absolute error by the measurement. For 10.0 ± 0.5 mm, the relative error is 0.5/10.0 or 0.05 (or 5%). In the case of estimating the error on an artifact density, we would usually measure an area on the ground and count all the artifacts in it. If we were to count 100 artifacts in an area of 20 m^2, it is easy to estimate our density as 100/20 or 5 artifacts/m^2. But we might want to account for the possibility that the count has some error (something we could check by having several people count the same area and see how much they vary); let's say that the estimated error is ±10 artifacts. That makes the relative error 10/100 or 0.1. If we also have a relative error on the measurement of area of, say ±0.03 (determined by propagating the errors on the length and width), our total relative error on the density would be the square root of $(0.1^2 + 0.03^2)$ or about 0.104. We now multiply this by the density to get the absolute error: 0.104 * 5 artifacts/m^2, or 0.52. Consequently, we should report the density as 5.0 ± 0.5 artifacts/m^2. In this case, the error is not noticeably larger than that from the counting error alone, because our estimated error on area was so small, but, in some cases, it can make a substantial difference.

1.3 Confidence and Validity in Measurements

Data also vary in the degree to which we can place confidence in them. For example, we are fairly likely to believe—at least within small margins of error—such claims as "this smoking pipe has a bore diameter of 3 mm" or "this sherd came from layer 6" but could be more skeptical of the claim that "feature 6 is a sweat lodge." At one time, archaeologists used to talk about the "ladder of inference" (Hawkes 1954) whereby some kinds of archaeological observations and interpretations were nearly certain, others fairly plausible, and still others highly speculative.

Without doing too much injustice to Hawkes's "ladder of inference," the idea is that we can infer from archaeological artifacts the techniques used to make them with a high degree of confidence and infer subsistence and economies with a fair degree of confidence. Inferences about social and political institutions, however, are considerably less certain. Differences in house size and the like are observations that may themselves be quite secure, at least in some contexts, but the inference that those differences are due to the presence of a chief, a medicine lodge, or a temple could be quite tenuous. To infer religious institutions, spiritual life, symbols or meanings "is the hardest inference of all" (Hawkes 1954: 162).

Hawkes's ladder is no longer very fashionable, but it does emphasize the point that some observations and inferences are more certain than others. Using the perspective of this book, this is a matter of the confidence we can place in claims or, to put it another way, how convincing they are. More generally, there are two kinds of observations that differ considerably in the degree of confidence we are likely to place in them.

Direct measurement involves fairly straightforward measurements, such as length or mass, where we can directly compare an object of interest with a standard scale. For example, we can measure the length of a projectile point by comparing it with tick-marks on a ruler or measure the mass of a coin by comparing it on a beam balance with the mass of one or more standard weights. We can only measure a quantity directly if we can make the measurement without having to measure some other quantity (Kyburg 1984: 90–112).

Indirect measurement involves measuring one phenomenon as a way of deriving a measure of some other concept. It is a type of measurement that is a lot more common than you might think. Even measurements of speed and temperature are indirect; we usually measure speed by first measuring distance and time and often measure temperature by measuring the height of a column of mercury (Kyburg 1984: 100, 113–42).

The simplest indirect measurements are ratios. For example, archaeologists never measure artifact densities directly; instead they count how many artifacts there are in some area of space (itself usually indirectly measured by multiplying length by width), and then divide artifact count by area to result in "sherds per square meter" or the like.

Often ratios make no reference to units, because those cancel out during division. So, the ratio of the frequencies of blades to flakes in a lithic assemblage (e.g., 1:1.4) or the proportion of Deverel-Rimbury pottery in a ceramic assemblage (e.g., 0.29 or 29%) is a unit-less indirect measurement ("percent" is not a unit, but simply shows that you have multiplied a proportion by 100).

In addition to the rather simple example of ratios, there are conceptually indirect measures that bring to mind Hawkes's "ladder of inference." There are many kinds of things that archaeologists would like to measure, but cannot measure directly, so they try to think of some kind of "proxy" measurement that might be related to the measure of interest. Some examples include:

- The number of people who inhabited a Tsimshian plank house
- The wealth of a Bronze Age household
- The social status of the occupant of a grave or tomb
- The degree of social or economic interaction between neighboring settlements
- The volume of traffic of obsidian between two sites in an exchange system
- The amount of deer meat in a Mesolithic diet
- The mesh size that a prehistoric farmer used to sieve seeds found in an assemblage of plant remains
- Variation in the magnitude of prehistoric mean annual rainfalls
- Conservatism in pottery decoration

These and many other observations that archaeologists would like to make are indirect because, short of a time machine, we cannot go back and measure them directly and have to depend on various lines of evidence that allow us to infer or estimate such quantities. To measure ancient population sizes, for example, archaeologists have proposed many different indirect measures based on site area, roofed settlement area (Fig. 1.3), number of hearths, number of burials per century, length of longhouse, amount of architectural rubble, area and density of sherd scatters, and even average volume of cooking pots multiplied by number of hearths (e.g., Berrey 2018; Duwe et al. 2016; Chamberlain 2006; Hassan 1981; Naroll 1962). Each of these measures is supposed to have some reasonably predictable relationship with a site's or house's ancient population size, but that relationship could be quite complex. For example, to base the estimate of a community's population on the number of burials in a cemetery, you need to make assumptions about average life expectancy, whether all or only some people from the community where buried there, and the duration of the cemetery's use. With indirect measurements such as these, it would be

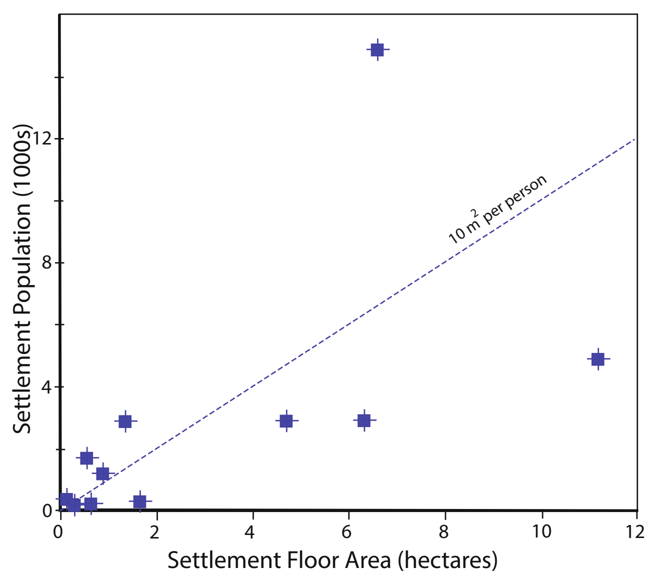


Fig. 1.3 Scatterplot of the relationship of total floor area to population size in an ethnographic sample of largest settlements (Modified from Naroll 1962). The graph omits the largest site (Cuzco at 16.7 ha and 200,000 people) and has linear, rather than logarithmic scales. The dashed diagonal indicates an alleged relationship of 10 m^2 of floor area per person but note how poorly the data fit the regression line, which is heavily influenced by the three largest sites, especially Cuzco. Small error bars have been added just to illustrate how to show estimated measurement errors on area and population statistics

reasonable to expect a rather large margin of error, and correspondingly lower confidence than we would place in the direct measurements on which we based them.

Even if we could measure site size very precisely and accurately, for example, estimates of population based on site area could have errors due to the following:

- There is considerable variation in the ethnographic examples we use as a basis for estimating an average number of people per unit area.
- There may be errors in the selection of these ethnographic examples or inclusion of completely inappropriate examples.
- Possibly not all of the site was occupied at the same time, so the estimated number of site occupants is biased upward.
- Possibly not all of the site was used for settlement, with some parts used for gardens, industrial areas, or livestock. This also results in upward bias.
- Possibly the site is atypical, with little resemblance to the ethnographic cases or other sites of the same culture. For example, it might be a fortified site in which people crowded during time of war.
- The relationship between site size and the number of people occupying sites may simply not be very strong or other factors, such as the length of time people anticipate they will occupy the site ("anticipated mobility") may be more strongly correlated (Kent 1991: 39).

Errors in direct measures are compounded when you apply arithmetic operations such as division and multiplication. Consequently, you can expect any operations you use to make indirect measures to magnify the measurement errors you originally had.

1.3.1 Validity

Validity is the extent to which our choice of measure (for that matter, also our choice of other aspects of a research design) is a good reflection of the thing we really want to measure, which makes it particularly relevant to indirect measurement. For example, one might use bone density as an indirect measure of the probability of bone preservation (see Chap. 15), in which case we should be concerned with how valid this is. Some indirect measures are clearly more reasonable than others, and having a high degree of reliability in measurement does not guarantee that measurements will be valid ones.

Social scientists typically distinguish among several kinds of validity. The simplest and weakest is **face validity**, which is just an assessment of whether or not the measure seems, subjectively, to make sense. Sometimes thinking about face validity is all we have to do to eliminate some kinds of measures because, without doing any formal analysis, we can see that they do not make sense. For example, it should be clear that using only the number of bones belonging to various animal species found at a site as a way to measure those animals' relative contributions to the diet of the people who once lived there would not be valid, because the bones of large animals would have carried a lot more meat than those of small animals. This is a problem to which we return in Chaps. 7 and 15.

Content validity has to do with whether our way of making an indirect measure covers all the relevant aspects of the problem. A valid measure of the wealth of a Bronze Age household, for example, probably should not be based only on the number of metal artifacts found in a house building or in graves. Not only are metal artifacts typically rare and subject to recycling and a host of other complicating site-formation processes, they are only one aspect of wealth. A more valid assessment of wealth differences among households or graves might take into account a range of evidence, such as clues to household diet and health status, the size and number of storage facilities, the size and elaboration of house structures or graves, the presence of non-metallic luxury items or imported items, and potentially others.

Construct validity is the degree to which any measure actually measures what it purports to measure. It is a combination of convergent validity (the degree to which two measures that theoretically should be associated or correlated, are associated) and discriminant validity (the degree to which two measures that should not be related, are not). For example, we might expect a measure of wealth based on "expensive" artifacts to be correlated with measures of health and life expectancy based on skeletal evidence. Meanwhile, we would hope that it is not correlated with things like research intensity (e.g., proportion of excavation), observer training, or conditions of preservation. If there are associations with these kinds of variables, it may be that something like poor preservation is a better explanation of variation than a hypothesis about wealth differences, a major threat to the validity of the latter.

Criterion-related validity involves assessing our measures against some criterion to see how well they perform, and thus is stronger than face validity. One way to do this is through **predictive validity**; for example, we could take our preferred method for estimating the number of people who would have occupied a site and compare the results with those from applying competing methods for measuring the same thing. Consistency between two methods would be somewhat reassuring, but it is still possible that both are wrong. Consequently, it is better to apply our measure to some relevant ethnographic cases (although not the ones we used to create our measure—see **independence**,

p. 131) to see how well it predicts the known population sizes for those cases. Sometimes, we can do this by checking for the strength of the statistical correlation between two competing proxies or between predicted and actual values. Another criterion-related type of validity, **concurrent validity**, is particularly useful when we are trying to classify or group things (see Chap. 3). To demonstrate this type of validity, we use tests to check how well our measures and methods allow us to assign observations or artifacts to relevant categories or distinguish groups that make sense. For example, we might do statistical tests with whole vessels to see how well measures we designed to assign sherds (vessel fragments) to various vessel shapes (bowl, jar, etc.) actually perform. Although this also has a predictive aspect, the more important criterion for concurrent validity is how well we can discriminate between groups even when those groups are fairly similar.

Validity is also a topic that comes up in research design, so we will return to it in Chap. 6.

1.4 Quality Assurance and Quality Control

Quality Assurance (or QA) involves policies, procedures, manuals, standards, and systems that one can use to ensure and improve the quality of products or services and consumers' satisfaction with them (Arora 1998; Hughes and Williams 1995; Schlickman 2003; Willborn 1989). It is a term you will find most commonly in industrial and commercial contexts, but it is possible to apply it in archaeology as well (e.g., Banning et al. 2016; Hunt and Sadr 2014; Nims and Butler 2017; Whittaker et al. 1998; Wolverton 2013).

In an archaeological context, the "product" would be the data we produce with our observations and measurements, or the conclusions we draw from the data, and it is our job to create policies, procedures, and standards that help us ensure that the data are sufficiently accurate, precise, reliable and valid for the purposes we have set for them. Our "consumers" are other archaeologists and others, such as heritage managers and the public, who make use of our data, perhaps when they read one of our archaeological reports. We can examine data quality from two perspectives: **quality assessment**, which involves evaluating the quality of measurements made in the past and whether that has increased or not, and **quality control**, which involves ensuring that ongoing work meets acceptable quality standards.

Some of the aspects of QA that are particularly relevant to archaeology pertain to the more practical facets of research design (Chap. 6). In order to allow confidence in archaeological observations, especially when more than one person will make them or they need to be compared to data from another project, it is useful to have strict protocols and standards for how measurements will be made, how to reduce the impact of inter-observer differences, how much tolerance there will be on measurements, how to define various terms, how to apply a typology, how to enter data in a database, and so on. One way to do this is to create a lab manual that outlines procedures and sets standards for acceptable measurements.

1.4.1 Tolerance

Tolerance is something that comes up most frequently in Engineering and industry, where the focus is on setting limits on the acceptable variation in such things as machine parts. However, it can be relevant to archaeology too, because an important aspect of tolerance is how much measurement error is acceptable in a Quality Assurance framework.

For archaeological purposes, we might think of this as a target range of acceptable error, somewhat like goal posts or the space between the tails of a normal distribution (see Chap. 8), and closely related to the concept of reliability. Whether we are measuring stone tools or prospecting for sites in a landscape, we want not only to ensure that different observers get much the same results (reliability), but that the range of variability that is left is within this acceptable range.

Tolerance is also an aspect of deciding appropriate sample sizes. The effective question is, how large a sample do I need to make a decision about my data with a risk of error no larger than x? We return to this in Chap. 6.

1.4.2 Audits and Inter-Rater Reliability

QA also typically includes audits; that means having someone, preferably an expert, spot-check results. This typically means checks on inter-observer differences (or Inter-Rater Reliability, IRR) between the expert and one or more other "raters" in identifications or measurements, or even differences in measurements by the same analyst at different times (Intra-Rater Reliability, e.g., Lyman and VanPool 2009). A weakness, of course, is that the expert rater cannot be immune from error either, which is the main weakness of error rate.

A simple measure for inter-observer comparisons of nominal (typological) data is the **joint probability of agreement**, which is simply an estimate of the proportion of agreements among two observers, and thus something like the "flip side" of error rate. However, this is not a robust measure, as it does not take into account that agreement sometimes occurs just by chance. **Cohen's kappa** (for two observers) and **Fleiss's kappa** (for multiple observers) are more robust because they compensate for chance agreements (Cohen 1960; Fleiss 1971).

For comparing observers' measurements on ordinal, interval, and ratio scales, one can use paired and intra-class correlations (Lyman and VanPool 2009; Shrout and Fleiss 1979). However, the most versatile statistic for comparing measurements is **Krippendorf's alpha**, which assesses agreement among observers at any scale of measurement (see Hayes and Krippendorf 2007).

An aspect of QA that is particularly apt here is the field of **statistical quality control**. It is explicitly concerned with errors and just how much error we are willing to tolerate.

There are two main pillars of statistical quality control that are highly applicable to archaeological laboratory work.

One of these is the concept of **acceptability sampling**, which involves sampling a certain number of observations to be checked or audited by an expert (e.g., Almagro-Gorbea et al. 2002). If the auditor finds that the sample has been correctly measured and described, it may be possible to relax the frequency of the checks. If, on the other hand, the auditor finds data that do not fit within tolerances, the whole batch of observations is rejected and has to be reanalyzed, and the frequency of checking will increase.

Closely related to acceptability sampling is the **control chart** (Fig. 5.29). This allows us to assess visually how well sets of measurements conform to our standards over a long period of time, thus making it easy to identify occasions when there is a sudden change in quality (Montgomery 2009). We then attempt to explain the change: was there a change in lab personnel or a new training session? Did we introduce a new instrument or method that lab volunteers found confusing? Finally, we can try to fix any problems that a spike in the control chart might suggest, sometimes through additional training, sometimes by improvements to manuals, equipment, or procedures. We will return to control charts in examples in Chaps. 5 and 12.

1.5 How Much Error Is Reasonable?

At the end of the day, some errors are more worrying than others and some are not worth any worry at all. It is important to ask yourself just how much measurement precision is warranted, given your research questions and those of your likely audience. Will it really matter or alter our interpretation if our estimate of site size is out by 3%, by 5% or by 10%? Will our ultimate use of the data involve lumping it into ordinal categories? In most cases, we will find that rounding off measurements to only two or three significant digits is a lot more honest than implying high degrees of precision and accuracy that are unnecessary and often quite meaningless. As mentioned above, this is a matter of tolerance.

On the other hand, it is not acceptable to act as though the errors do not exist. An emphasis throughout this book is honest reporting of results, including their likely errors, while trying to minimize those errors where they could have unfortunate impacts on our interpretations, as well as recognizing situations where the errors really are not that important. This distinction brings us back to validity, because our main concern is whether or not such errors are likely to lead to a bad decision (such as bulldozing a site that should have been preserved) or a bad interpretation (such as rejecting a correct hypothesis about the function of a site or inferring an incorrect age for a site's abandonment). Errors are inevitable. What matters most is how we deal with them.

1.6 Summary

- Data are not "given" but result from questions we ask and decisions we make in trying to answer them
- We measure data in different ways that we can characterize as nominal, ordinal, interval and ratio scales
- Data inevitably have errors and uncertainty, but we can estimate these and try to minimize the most critical ones
- The validity of measurements and the inferences we make from them depends on both common sense and careful analysis to determine whether we are really measuring what we think we are measuring
- Quality Assurance and Statistical Quality Control are tools to help us ensure better validity in research

References Cited

Almagro-Gorbea, M., Alonso, P., Benito, J. E., Martin, A. M., & Valencia, J. L. (2002). Statistical quality control in archaeological survey. *Archaeological Prospection, 9*(2), 45–53.

Arora, S. C. (1998). *Applying ISO 9000 quality management systems*. Geneva: International Trade Centre.

Banning, E. B., Hawkins, A., Stewart, S. T., Hitchings, P., & Edwards, S. (2016). Quality Assurance in archaeological survey. *Journal of Archaeological Method and Theory, 24*(2), 466–488.

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data*. New York: Wiley.

Berrey, C. A. (2018). Making absolute population estimates in the intermediate area using the area and density of ceramic sherd scatters: An application of regression analysis. *Journal of Archaeological Science, 97*, 147–158.

Binford, L. R. (1964). A consideration of archaeological research design. *American Antiquity, 29*(4), 425–441.

Chamberlain, A. (2006). *Demography in archaeology*. Cambridge: Cambridge University Press.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

Daniels, S. G. H. (1972). Research design models. In D. Clarke (Ed.), *Models in archaeology* (pp. 201–229). London: Methuen.

de Waal, C. (2016). Charles Sanders Peirce and the abduction of Einstein: On the comprehensibility of the world. *History and Philosophy of Physics*. arXiv:1610.00132 [physics.hist-ph].

Duwe, S., Eiselt, B. S., Darling, J. A., Willis, M. D., & Walker, C. (2016). The pueblo decomposition model: A method for quantifying architectural rubble to estimate population size. *Journal of Archaeological Science, 65*, 20–31.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin, 76*(5), 378.

Hand, D. J. (1997). *Construction and assessment of classification rules*. New York: Wiley.

Hanson, N. (1958). *Patterns of discovery*. Cambridge: Cambridge University Press.

Hassan, F. A. (1981). *Demographic archaeology*. New York: Academic.

Hawkes, C. (1954). Archaeological theory and method: Some suggestions from the Old World. *American Anthropologist, 56*(2), 155–168.

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77–89.

Heilen, M., & Altschul, J. H. (2013). The accuracy and adequacy of in-field artifact analysis. *Advances in Archaeological Practice, 1*, 121–138.

Hughes, T., & Williams, T. (1995). *Quality Assurance. A framework to build on* (2nd ed.). Oxford: Blackwell Science.

Hunt, T., & Sadr, K. (2014). Inter-analyst variability in identification and classification of pre-colonial stone-walled structures using Google Earth, southern Gauteng, South Africa. *The South African Archaeological Bulletin, 69*(199), 87–95.

Jones, A. M. (2002). *Archaeological theory and scientific practice: Topics in contemporary archaeology*. Cambridge: Cambridge University Press.

Kehoe, A. B., & Nelson, S. (1990). Introduction. In S. M. Nelson & A. B. Kehoe (Eds.), *Powers of observation: Alternative views in archeology* (pp. 1–4). Washington: American Anthropological Association.

Kent, S. (1991). The relationship between mobility strategies and site structure. In E. M. Kroll & T. D. Price (Eds.), *The interpretation of archaeological spatial patterning* (pp. 33–59). New York: Plenum Press.

Kyburg, H. E. (1984). *Theory and measurement*. Cambridge: Cambridge University Press.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Lyman, R. L., & VanPool, T. (2009). Metric data in archaeology: A study of intra-analyst and inter-analyst variation. *American Antiquity, 74*(3), 485–504.

Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed.). New York: Wiley.

Naroll, R. (1962). Floor area and settlement populations. *American Antiquity, 27*, 587–589.

Nims, R., & Butler, V. L. (2017). Assessing reproducibility in faunal analysis using blind tests: A case study from northwestern North America. *Journal of Archaeological Science: Reports, 11*, 750–761.

Peralta, M. (2012). *Propagation of errors: How to mathematically predict measurement errors*. Seattle: CreateSpace.

Pirsig, R. M. (1974). *Zen and the art of motorcycle maintenance: An enquiry into values*. New York: Morrow.

Proulx, B. B. (2013). Archaeological site looting in ‘Glocal’ perspective: Nature, scope, and frequency. *American Journal of Archaeology, 117*(1), 111–125.

Schlickman, J. (2003). *ISO 9001:2000 quality management system design*. Boston: Artech House.

Shanks, M. (1992). *Experiencing the past: On the character of archaeology*. London: Routledge.

Shanks, M., & Tilley, C. (1987). *Re-constructing archaeology. Theory and practice*. Cambridge: Cambridge University Press.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*(2), 420.

Sifniotis, M., Watten, P., Mania, K., & White, M. (2007). Influencing factors on the visualization of archaeological uncertainty. In *VAST'07, proceedings of the 8th international conference on virtual reality, archaeology and intelligent cultural heritage* (pp. 79–85). Aire-la-Ville: Eurographics Association.

Thomas, J. (2013). *The birth of Neolithic Britain: An interpretive account*. Oxford: Oxford University Press.

Tilley, C. (1994). *A phenomenology of landscape: Places, paths and monuments*. London: Bloomsbury Academic Press.

Whittaker, J. C., Caulkins, D., & Kamp, K. A. (1998). Evaluating consistency in typology and classification. *Journal of Archaeological Method and Theory, 5*(2), 129–164.

Willborn, W. (1989). *Quality management system, a planning and auditing guide*. New York: Industrial Press.

Wolverton, S. (2013). Data quality in zooarchaeological faunal identification. *Journal of Archaeological Method and Theory, 20*, 381–396.

2 Summarizing Data: Descriptive Statistics

Although a picture or diagram of some kind may very often give us the summarised information we want, to do anything more with that summarised information it needs to be in the numerical form of descriptive statistics so that it can be further manipulated.

(Shennan 1997: 34–35).

While we can summarize data in tables and graphs (see Chap. 5), sometimes it is useful to use statistical statements that summarize a lot of data in one or two numbers. This can sometimes make it easier to recognize patterns and is also the basis for statistical analyses (see Chap. 8).

Verbal summaries have a long history in archaeology. An archaeologist may describe "typical" examples, such as the most common kinds of pottery in a site along with noteworthy exceptions that give some impression of variation in the data, for example. Some verbal summaries can be richer, and are often more interesting, than numerical summaries, and they have the advantage that they can convey some of the researcher's thought processes and goals (Hodder 1989). However, they are not amenable to accurate comparisons of data sets, whether using statistical methods or not. They have an important role in archaeology but should not be the exclusive means of archaeological reporting.

Numerical summaries of interval or ratio-scale data are what archaeologists usually mean by **descriptive statistics**. These are measures intended to sum up the "typical" or "central" characteristics of the data (central tendency) or the amount of "spread" in the data. For nominal-scale data, we can summarize data by relative abundance (**proportions** or percentages). However, some other descriptive statistics are also applicable to nominal or ordinal scales. For more on archaeological descriptive statistics, see Drennan (2010: 17–36) and Shennan (1997: 5–20).

2.1 Central Tendency

The most common measure of central tendency for interval and ratio scales is the simple average, or **arithmetic mean**. For a sample (a subset of observations taken from a population, see sampling in Chap. 6), the statistical expression for averaging is,

$$\bar{x} = \frac{\sum x_i}{n}$$

where $\bar{x}$ (or bar-x) is the sample mean, x_i are observations of some kind on all the members of the sample (e.g., x could be the lengths of projectile points), and n is the number of members in the sample, or sample size. The "i" subscript on x means any individual x value. This simply means that we add up all the individual values and then divide by the number of observations. We could also write the numerator as $x_1 + x_2 + x_3, \ldots + x_n$, to indicate summing all the values from the first one to the n-th one, but Σx_i is a shortcut. For example, the mean of 5 cm, 5.7 cm, and 6.3 cm would be (5 + 5.7 + 7.3)/3 = 18/3 = 6 cm.

The mean is something like a "center of gravity" for a distribution of numbers and takes all those numbers into account. If you were to calculate the "deviations" from the mean for each observation ($x_i - \bar{x}$, the difference between each observation and the mean), the sum of these deviations would be 0, indicating a sort of "balance" around the mean.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_2

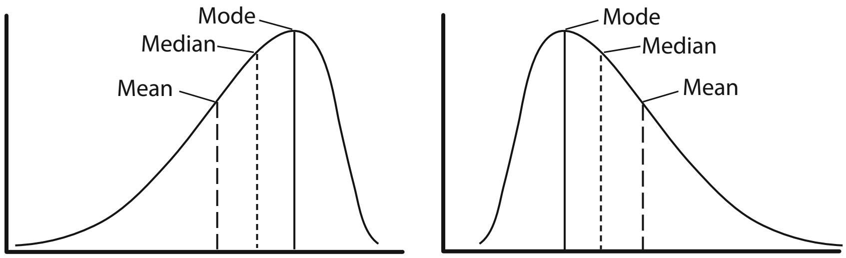


Fig. 2.1 The effect of symmetry and skewness on some common measures of central tendency. In a normal distribution, the sample mean, median and mode are equal and in the center of the curve. In skewed distributions, however, the sample mean deviates considerably from the median and mode because extreme values in the long tail pull it to left or right. In such cases, the mean does not provide a good sense of central tendency

However, the sample mean is very sensitive to extreme values — ones that deviate quite a lot from the more typical measures in the sample — and any strong deviations from a normal distribution (Fig. 2.1, see Chap. 8), such as a tendency to be skewed to the right (see p. 20). Consequently, statisticians consider it to be the least robust measure of central tendency (Hole 1980: 228–229).

While archaeologists still make considerable use of means, and they have some useful statistical properties (see Chap. 8), it is often better to use a more robust measure of central tendency. One of these is a truncated or **trimmed mean**. This involves discarding data from the upper and lower parts of the distribution to remove the impact of extremes. For a 5% trimmed mean, for example, we remove both the upper 5% and lower 5% of the values and calculate a mean on the basis of the more central 90% of the data that remain. 10% and 20% trimmed means are common in many fields. For some reason, most archaeologists do not use trimmed means, but they are a valuable alternative to the common mean, especially for skewed distributions (Fig. 2.1) or cases with suspected outliers.

Another alternative is the **median**. The median is simply the value that divides the observations into two groups, such that half are below, and half above the median. Thus, it is equivalent to the 50th percentile. Because the median only references how many observations are above and below, rather than how far away they are, it is not sensitive at all to extreme values. Technically, it is an extreme example of a trimmed mean, with all the data trimmed out except for the middle one or two values. The median is also applicable to data measured on ordinal scales, simply taking the middle value when all the data are ranked in order,

Another common measure of central tendency is the **mode**, which is applicable to abundances in nominal-scale and ordinal-scale data, as well as to interval and ratio measurements. It is simply the most common value in the data. However, because no two continuous values are truly identical, continuous interval and ratio data in typical samples must be grouped into equal intervals or "bins," as

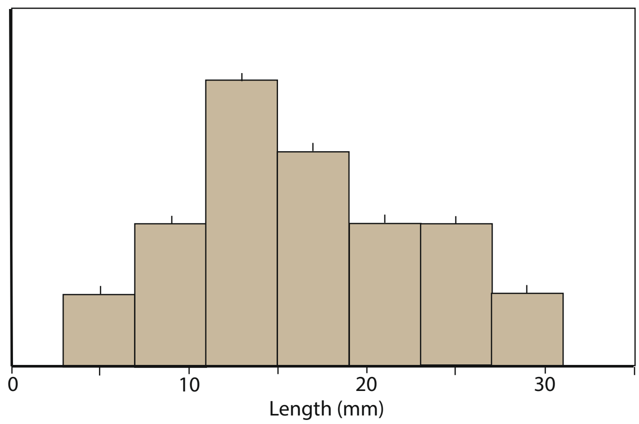


Fig. 2.2 A histogram showing how to calculate a mean from grouped data, as you might find in a publication. Here we would multiply the number of observations in each interval by the mid-point of that interval (indicated by ticks on top of the bars), sum those values, and then divide the sum by the number of observtions. In this example, the sum would be 5 + (2*9) + (4*13) + (3*17) + (2*21) + (2*25) + 29, so the mean would be 247/15 = 16.5 mm

in a histogram (see Chap. 5), so that the mode corresponds with the highest point on a graphed distribution.

It is also possible to measure central tendency in more than one dimension. For example, in two dimensions, as in the distribution of observations on a map, the sample mean is at the coordinates $(\bar{x}, \bar{y})$, called the **mean center**.

Sometimes you need to calculate central tendency, such as a mean or trimmed mean, from data that have already been summarized to some extent. In a graphed frequency distribution, for example (Fig. 2.2), we can just count how many observations are in each "bar" of the graph (the "frequencies"), multiply that by the middle x-value for the bar, and then sum these values, which gives us a pretty good estimate of $\sum x_i$. We can then divide that sum by the total number of observations across all the bars in the graph, which is equal to n, and obtain an estimate of the mean. We can do the same thing except to omit an equal number of the lowest and highest values to obtain a trimmed mean, while we could estimate the median as the value at the middle of the bar that contains the observation that splits the data into two groups, the lower $n/2$ and the higher $n/2$.

2.2 Dispersion

Aside from central tendency, it is generally useful to know how spread out the data are, the dispersion.

On an interval or ratio scale, the simplest measure of dispersion is the **range**. This is just the difference between the lowest and highest values. Consequently, its value depends on only two measurements, and ignores all the other data. Because these are also the two most extreme values in the data, this is not very desirable. Furthermore, the range tends to increase whenever we increase our sample size, so it is not very stable. Although there are some instances where range is really important — organisms die if the temperature or humidity in their environment goes outside a survivable range — in most cases, range is not a very good measure of dispersion in data.

In cases where we might use a median for central tendency, one appropriate measure of dispersion is the **interquartile range** (IQR). This is the range that encloses the middle 50% of the observations. In conjunction with the median, then, it divides the range of values into four groups, the bottom 25%, the 25% below the median, the 25% above the median, and the upper 25%. It is thus the range between the 25th and 75th percentiles.

Another measure of dispersion appropriate to medians is the **median absolute deviation** (Hoaglin et al. 1983: 404–414). It is a robust measure of dispersion and is simply the median of the absolute deviations from the median in a sample or population. What this means is that we subtract every observation from the median (the deviations) and find the median of these values. For symmetrically distributed data, it yields the same result as IQR. It is a useful and, in archaeology, underused statistic (e.g., Leys et al. 2013).

For cases where we use a mean for central tendency, it would make sense to use a measure of dispersion that involved the distances of all the data values away from the mean. As we have seen above, the average of all these distances is 0, so average deviation would not work. We could average the absolute values of the distances, and this is called the **mean absolute deviation** or just "mean deviation." Although some people like the simplicity of this measure, archaeologists have rarely used it and, unfortunately, its common abbreviation as "MAD" is unhelpful as it creates ambiguity with the median absolute deviation. The mean deviation could be used around a mean, a mode, or a median, typically yielding different results, which makes the ambiguity even worse. Another reason its use is uncommon is that, for samples, it is a biased estimator of dispersion in the population. On the other hand, its ratio to the standard deviation can serve as a test of normality (Geary 1935).

The most common measures of dispersion around a mean are **variance** (s^2) and **standard deviation** (s). Sample variance is an average of the squared deviations around the mean:

$$s^2 = \frac{\sum (\bar{x} - x_i)^2}{n - 1}$$

In other words, you measure the difference between each value and the mean, square these, sum them, and divide by the size of the sample minus 1. Using the same simple example as for the mean, the sum in the numerator would be $(6-5)^2+(6-5.7)^2+(6-7.3)^2=1^2+1.3^2+1.3^2=1+1.69+1.69=4.38$. Dividing this sum by $(3-1)$ or 2 gives us a variance of about 2.2 cm^2 (2 significant digits).

As the units of variance are not the same as those for the mean — if the mean is in cm, the variance is in cm^2 — it is often more convenient to take the square root of the variance to obtain s, which we call the **standard deviation**. In this case, that would be $\sqrt{2.2}$, or about 1.5 cm.

Sometimes it is useful to standardize the standard deviation relative to the mean. This is called the **coefficient of variation** (CV, or relative standard deviation, RSD), which is simple to calculate by dividing the standard deviation by the mean and often expressing the result as a percentage. Because it requires this division, it is only applicable to ratio scales, and when the mean cannot be zero.

Just as we can have trimmed means, we can also have **trimmed standard deviations**. As you would expect, this is just the same as calculating a regular standard deviation, except that you omit the same lowest and highest values as for the trimmed mean.

Yet another measure of dispersion, specifically for sampling distributions, is the Standard Error (SE). It is based on the standard deviation but takes sample size into account:

$$SE = \frac{s}{\sqrt{n}}$$

It is always smaller than the standard deviation unless $n = 1$ and has very useful statistical properties (see Chap. 8). Dividing SE by the mean results in **relative standard error** (RSE), typically expressed as a percentage. Like CV, it is only applicable to ratio-scale data.

Finally, one useful measure that is similar to some of these, but specifically designed to measure the differences between replicated measurements in samples, is **technical error of measurement** (TEM):

$$TEM = \sqrt{\frac{\sum d_i^2}{2n}}$$

where d_i is the difference between two independent measurements of some quantity, and n is the number of comparisons. Archaeologists have sometimes used TEM, or its relative equivalent, to study intra- or inter-observer differences in measurement (e.g., Lyman and VanPool 2009). For example, if two archaeologists made independent measurements on the same five projectile points, and the differences between their measurements of thickness (d_i) were 0.4, 0.2, 0, 0.3, and 0.5 mm, we would square these differences to obtain 0.16, 0.04, 0, 0.09 and 0.25, sum these to yield 0.54, divide that by (2 * 5) to get 0.054, and then take the square root to obtain a TEM of 0.73 mm.

2.3 Skewness and Kurtosis

Skewness has already appeared in the discussion above. It is also something we can measure. Skewness is the degree of asymmetry in an interval- or ratio-scale distribution relative to its mean. There are many different ways of measuring skewness, including Pearson's first skewness coefficient, which is simply the difference between the mode and mean divided by the standard deviation, and Pearson's second skewness coefficient, which is the difference between the median and mean divided by the standard deviation.

Kurtosis has to do with the prevalence of outliers in the tails of an interval- or ratio-scale distribution. Distributions with high kurtosis have more extreme values than would a normal "bell-curve" distribution (see Chap. 8), so that they have relatively large "tails," while those with low kurtosis have unusually few outlying values. The most common measure of kurtosis is Pearson's fourth moment, in which a normal distribution has a kurtosis value of 3.

2.4 When Are Means Meaningless?

The sample mean is a mainstay of archaeology, even though it is not a statistically robust measure of central tendency. As already noted, among its problems are extreme values and skewness in data.

In some cases, furthermore, the mean depends far too much on the way we take measurements. A lot of the non-random error found in archaeological averages is due to this aspect of artifact densities and average artifact size. These depend a great deal on the *smallest* artifacts we include in our measurements. If one archaeologist only measures artifacts greater than 1 cm in maximum dimension while another ignores artifacts less than 2 cm in maximum dimension, they will arrive at very different average sizes *even if they are analyzing exactly the same set of artifacts*. Note that this is an example of low predictive validity, since archaeologists using different cut-offs at the low end will always end up with different estimates of the mean.

Since some artifact fragments will be even less than 1 mm in size, the lower boundary of what an analyst considers important or measurable is arbitrary and typically based on practical rather than theoretical considerations. Consequently, we should accompany any statement about mean artifact size with clear statements about how the measurements were made and what was the lower cut-off.

The same thing happens when we count artifacts. The number of artifacts we count from a particular context depends on how big they have to be to be counted. Again, anyone who counts all artifacts 1 cm or larger in maximum dimension will obtain a higher count (probably a much higher count) than someone who only counts ones 2 cm or larger. Since artifact densities are really just the average number of artifacts per some spatial unit, they are also very sensitive to this problem. Again, we need to be clear about what did and did not qualify as a countable artifact.

We also need to be mindful of other ways that averages can suffer from bias. In analyses that depend on patterns in site size, for example, we need to keep in mind that surveys are more likely to discover large sites than small ones. As a result, means and indeed other statistics that are sensitive to underrepresentation of small sites are biased.

As it turns out, the sizes and densities of things, including sites and artifacts, have *fractal* properties, a topic to which we will turn in Chap. 7. In some instances, it may make more sense to measure their **fractal dimension** instead of their density or mean size.

2.5 Summaries of Nominal and Ordinal Distributions

When we collect data with nominal and ordinal scales, as noted above, the natural thing is to count how many observations fall into each category, a process called enumeration.

Archaeologists often present enumerated data in the form of tables. For example, they may use tables to summarize the numbers of various types of stone tool or of various taxa of animals among faunal remains. Typically, the tables include totals and subtotals for various groupings of categories (Table 2.1).

Table 2.1 Distributions of lithics by tool categories from two stratigraphic phases at a site, by enumeration (n) and percentage. Two percentages are highlighted to draw attention to a major difference

	Layer 1		Layer 2	
Tool category	n	%	n	%
Burin	161	**49**	67	**27**
Endscraper	66	20	75	30
Notch	43	13	51	21
Microlith	39	12	31	13
Other	17	5	24	10
Total	326	100	248	100

Table 2.2 Example of a distribution of skeletal elements on an ordinal scale of preservation, with a median in the category for approximately 50% preservation because there are 207 bones and fragments in total, and the 103rd and 104th ones would be in the ~50% category

Small unidentifiable fragments	~ 20% of element	~50% of element	~75% of element	Complete element
64	22	47	41	33

However, it is difficult to compare the raw counts meaningfully, even if we standardize the definition of countable things, because of differences in sample size. Consequently, it is useful to standardize them to facilitate comparison. The parameters and statistics for this purpose are **proportions**, sometimes presented as percentages by multiplying by 100.

A proportion is just the ratio of the total count for a particular category to the total for all categories combined. To estimate the proportions in samples, we use,

$$p_i = \frac{m_i}{m}$$

with m_i being the number of items in category i in the sample, and m being the total number of items of all categories in the sample. The standard deviation of a sample proportion is,

$$s = \sqrt{p(1-p)}$$

so that the Standard Error is,

$$SE_p = \sqrt{\frac{p(1-p)}{m}}$$

with m again as the total of all the countable items in the sample.

Some of the measures that are applicable to interval and ratio scales are also applicable to nominal or ordinal scales, or both. In a table like that in Table 2.1, the mode is just the category with the most frequent observations (burins). In an ordinal scale, the mode would similarly be the category most frequently observed, but you can also identify the median — the category that divides the distribution in half (Table 2.2).

There are also other ways for us to summarize enumerated data on nominal and ordinal scales.

One of these is **diversity**, which has to do with how spread out the data are among categories. We will return to this type of summary in Chap. 7.

2.6 Summary

- Verbal summaries of archaeological results are very useful, but not very amenable to detailed comparisons
- Descriptive statistics offer ways to summarize numerically both central tendency — where the most common values lie — and dispersion — how "clumped" or spread out they are
- Skewness describes how much the distribution departs from a symmetrical pattern by having outliers far to the left or right of the more "typical" values
- For data that consists of counts of things in categories (enumeration of things in nominal or ordinal scales), we can standardize their comparisons by using proportions (or percentages), while we can also find medians for ordinal scales

References Cited

Drennan, R. D. (2010). *Statistics for archaeologists: A common sense approach*. New York: Plenum Press.

Geary, R. C. (1935). The ratio of the mean deviation to the standard deviation as a test of normality. *Biometrika, 27*(3/4), 310–332.

Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. New York: Wiley.

Hodder, I. (1989). Writing archaeology: The example of site reports in context. *Antiquity, 63*, 268–276.

Hole, B. L. (1980). Sampling in archaeology: A critique. *Annual Review of Anthropology, 9*, 217–234.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology, 49*(4), 764–766.

Lyman, R. L., & VanPool, T. (2009). Metric data in archaeology: A study of intra-analyst and inter-analyst variation. *American Antiquity, 74*(3), 485–504.

Shennan, S. (1997). *Quantifying archaeology* (2nd ed.). Iowa City: University of Iowa Press.

3 Systematics: Classification and Grouping

Poorly formulated typologies, human errors in classification, and theoretical biases may disrupt our ability to understand the typologies of others, to evaluate their interpretations, or even to be sure that our own are free of systematic errors

(Whittaker et al. 1998).

. . . artefact-types are conceived in terms of detailed sets of similarities between numbers of artefacts such that the degree of similarity between artefacts within the type group is greater than any similarity between artefacts in separate type groups

(Clarke 1968: 189).

Any useful comparison of archaeological data needs to have units with which to order the data so that we can find similarities and differences. The ways to formulate and structure such units constitute the methods of **systematics**, and a system of categories of being or concepts, and the relationships among them, is an **ontology**. Archaeologists typically describe most or all methods of systematics as "classification" or "typology," using these terms almost interchangeably, but they can also result in non-classificatory arrangements: groups or clusters. Ontology can more broadly concern how entities come into being but, in information science, it tends to focus on representations, definitions, and relations among concepts or data (Niccolucci et al. 2015).

Dunnell (1971) emphasizes strong intellectual distinctions between classification and grouping that are worth preserving, even though many archaeologists do not bother with this distinction. Conventional classifications, as developed further below, involve rules that must be satisfied for any entity to belong to a class. By contrast, grouping does not have this restriction, but instead takes advantage of "polythetic" similarities as found in the philosophy of Ludwig Wittgenstein and in some of the natural sciences (Needham 1975; Sneath 1962). This leaves conventional classification as "monothetic." It is also useful to distinguish classifications generally from particular kinds of classifications that we might call typologies.

This chapter introduces a classification of the tools that archaeologists use to order their data through systematics. It also provides a foundation for the definition of entities and attributes in archaeological databases (Chap. 4).

3.1 Systematics

Systematics operates on the nominal (or sometimes ordinal) scale of measurement. Our need for systematics stems from the fact that everything on which we would like to make observations is in some sense unique: each thing has its own particular combination of values on an infinite number of characteristics or **attributes**. When we try to carry out an analysis, however, it is not very useful to treat each item as unique or, alternatively, to clump everything into a single group. Categories or **classes**, and **groups** with members that share some attributes or are similar in some way help us to make sense out of data, not to mention saving analysis time and publication costs. Being able to group observations or assign them to categories is a necessary step in the intellectual process of making comparisons. If we literally treated every artifact, burial, or site as a unique observation, we would have no obvious way to find commonalities between artifact assemblages (which are themselves groups in any case). At the same time, to say that two artifacts belong to the same class, group, or type does not mean that they are identical in all respects, only that they share certain characteristics that a particular classification or grouping protocol highlights. The ontological position that these attribute-based categories and groups are "real" and ideals for which individual examples are only imperfect examples is called essentialism, and most modern archaeologists prefer to avoid essentialist characterizations, while recognizing some of the challenges this poses.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_3

One of the characteristics of systematics is that, philosophically speaking, the resulting ontological entities and relations are completely arbitrary. That does not mean that we do not have good reasons for organizing data in the way we do, but recognizes that the number of possible ways to define categories or group information is infinite, so we choose one way among many. Our research design may guide us to design systematics in particular ways—for example, a research goal to elucidate evolutionary relationships among living things leads us to use a taxonomic classification that at least mimics their evolutionary histories—but, in principle, no one arrangement is any better than any other. Only our theoretical orientations, preconceptions, research goals, methods, and plans for collaboration with others lead us to select one arrangement over another.

Archaeologists also tend to use words like classification and typology loosely and imprecisely. Even mathematicians are far from consistent in their use of terms for systematics. In this book, I follow Dunnell (1971) in distinguishing classification from grouping and I expand on Gardin's (1980) distinctions by treating typologies as special cases of classifications and groupings. While there are other ways of describing and distinguishing archaeological units of analysis (e.g. Adams and Adams 1991), Dunnell's usage in *Systematics in Prehistory* provides a clear and consistent set of terms and concepts.

Classification is the intellectual process of assigning items or instances, either real or imagined, to preconceived general categories, much as though we were putting objects into boxes or trays, so that they acquire an identity. Adams and Adams (1991) would call this process "sorting" when it involves classifying actual things, such as artifacts. Hand (1997) would call it "supervised classification." For each category, there is a rule or set of rules to determine whether any item belongs or does not belong to that category and, if we define the rules carefully, there is no ambiguity about the category to which any particular item belongs. The rules or definitions state the conditions that are both *necessary* and *sufficient* for assignment to that category: failure to meet even one of these conditions would disqualify it. A classification is an abstract arrangement with which we conceptualize the categories, or **classes**, to which we are assigning items, that is, with which we create the units of a nominal scale.

Grouping is a very different intellectual process. In this case, there is no pre-arranged system for ordering the phenomena of interest independent of the phenomena themselves and the kinds of attributes we have decided to observe. Grouping may be as simple as depositing a collection of artifacts onto a table and moving them around into piles in such a way that items in the same pile seem more similar to each other than they are to ones in other piles. Without rules on how to measure this similarity, different researchers are likely to create quite different piles. The important thing, philosophically, is that the starting-point is not an abstract model of how to conceptualize the items, but rather an actual collection of items. You might think of it as a bottom-up instead of top-down approach. You can have classes that have no members, but you'll never have groups without members. Hand (1997) would describe grouping as "unsupervised classification" or "pattern recognition."

Both classification and grouping belong to what Dunnell (1971: 43) calls arrangement—a procedure that orders data into units, an organizing device. An important point is that all arrangements are arbitrary; just as the number of attributes with which we might describe a particular object is infinite, the number of possible ways we could classify or group objects is also infinite. Although we routinely have theoretical justification for preferring one arrangement over another, from the formal standpoint one arrangement is just as real and valid as any other.

One reason that the difference between classification and grouping is important is that it affects our research designs. If our research design calls for us to anticipate the kinds of data we will collect and how we will collect them, as it often does, classification has to be part of that design, at least in its early stages. If we are doing exploratory analysis of a pre-existing collection, or working on data already sorted by classification methods, but only very generally, we might use grouping methods to get insights that might lead to the formation of a new hypothesis or to evaluate a hypothesis that the original research design did not anticipate. Sometimes, that grouping exercise can lead to a new classification. Typically, archaeologists use both classification and grouping at various stages of their research.

Let us examine the differences between grouping and classification in detail (Table 3.1).

Table 3.1 Characteristics of classification and grouping

Classification	Grouping
Abstract	Concrete
Consists of classes	Consists of things
Classes are mutually exclusive	Groups can intersect or overlap
Classes are defined	Groups are described
Assigning something to a class does not affect its definition	Adding or subtracting things from a group changes the group's description
Based on rules	Based on tendencies, lists, or boundaries
Rules specify conditions that are both necessary and sufficient for membership in a class	No condition is either necessary or sufficient for membership in a group
Items are assigned to pre-existing classes	Groups are formed from pre-existing items
Classes can have no members	There can be no group without members

Classifications are abstract. Any classification consists of a number of categories, or **classes**, each defined by one or more criteria, that exist outside the objects to be classified. Classifications are independent of the things they classify. Whether or not any artifacts belong to a particular class, whether you add items or take some away, the rules that define that class do not change.

The classes provide values on a nominal scale. To measure any item on this scale, we compare it with the criteria that define each class until we find a match. More formally, the definition of each class states the conditions that are both *necessary* and *sufficient* for membership in the class. What that means is that no item can be a member unless it meets all the criteria, while no item can be omitted as long as it meets those criteria. Consequently, classifications need to be formulated in such a way that the classes are both exhaustive (anticipate all possible items) and mutually exclusive; no artifact, site, or landscape element should satisfy the conditions for more than one class.[1] Ideally, this ensures that there is no ambiguity about the class to which it belongs (but see misclassification, and "fuzzy classification," below). Because classifications have definitions or rules for membership, we can also say that classification is **definitive** (Niccolucci et al. 2015: 91).

Grouping methods, by contrast, start with things, not rules. These things could be real or imaginary, directly in front of us or hypothetical. Instead of being definitive, grouping is descriptive, because we describe groups or clusters of things. We can describe them by enumerating their members, by stating their boundaries in space or time, or by summarizing the characteristics of each group statistically. We might describe a group of people, for example, by simply enumerating their names and addresses. Archaeologists routinely group artifacts by stating their boundaries within a site or survey unit (artifacts found in a pit or building, artifacts from Stratum II at Site 15), and we can in this case include "hypothetical" artifacts from spaces in the site that we haven't even excavated yet. We can also use chronological boundaries (e.g., artifacts from the Archaic period). A statistical description of a group of lithics might be that they tend to be 3–5 cm in length, have length/width ratios close to 1.7, and edge angles around 18°. Take note of the fact that these are not definitions, but just "tendencies." Just because the average length in the group might be 3.7 cm or length tends to be between 3 and 5 cm does not mean that there are no members of the group smaller or larger than these values, only that the majority of group members cluster in the way described.

One of the implications of the fact that grouping is descriptive is that, unlike classes whose definitions do not change, the description of each group changes, sometimes considerably, if you take away or add items (Dunnell 1971: 89). Adding a long flake to a group will increase the average length for that group, for example. Another is that groups cannot exist independently of their members. You can have empty classes, but you'll never have an empty group.

3.2 Classes of Classification and Grouping Arrangements

As with any phenomena, there are other ways to classify arrangements (e.g., Adams and Adams 1991: 216–228), but here we follow broadly the classification of Dunnell (1971: 44). He describes two major classes of classifications and two major classes of grouping methods (Fig. 3.1).

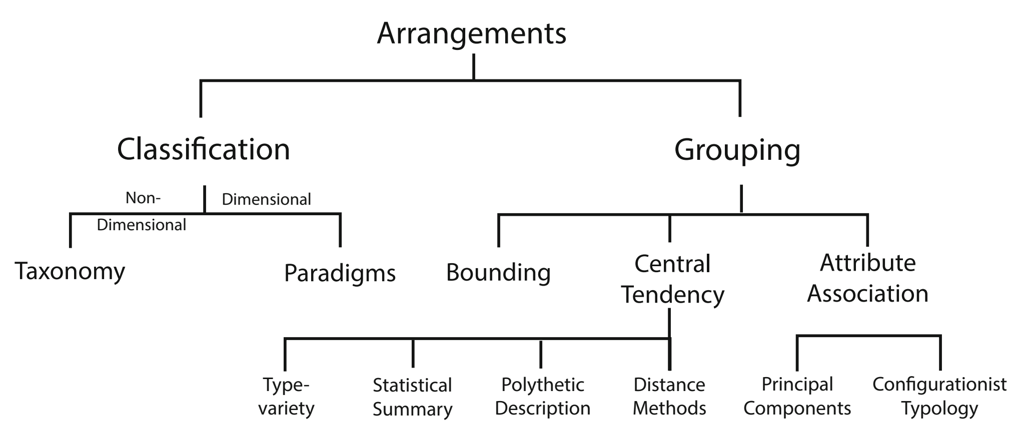


Fig. 3.1 A taxonomic classification of arrangements

[1] This means classes in the same classification system. Sometimes we may be using more than one classification in the same research project (e.g., a site might belong to a site type in a sites classification but also to a landscape type in an environmental classification), while the hierarchy that occurs in some kinds of classification allows (and actually requires) that items that belong to a class at the "bottom" of the hierarchy also belong to the larger classes to which that class belongs higher in the hierarchy.

Table 3.2 Paradigmatic classification of pottery rims by rim shape and stance angle

		Rim shape		
		Round	Flat	Bevelled
Stance	Inverted			
	Vertical			
	Everted			

3.2.1 Paradigmatic Classification

Paradigms operate by the intersection of nominal- or ordinal-scale dimensions (Table 3.2). That means that the classes on each scale are mutually exclusive, and the intersections of two or more of these scales define classes in much the same way as x, y, and z values (Cartesian coordinates) can define portions of space. Readers who have studied languages will be familiar with paradigms because they are used to classify verbs. In those cases, one dimension is "tense," often with classes for past, present and future. Another could be "gender," with feminine, masculine and perhaps neuter classes. Another is "number," with classes for singular, plural and perhaps dual. Yet another dimension is "person," with first person (associated with pronouns "I" or "we"), second person ("you") and third person (e.g., "she" or "they").

Paradigms have some important characteristics. Because they are dimensional, they are non-hierarchical and unweighted. All dimensions contribute equally to the classification, leading to a kind of symmetry. It is also possible, indeed quite likely, that a paradigmatic classification with many classes will have at least some "empty" classes—that is, some combination of values on the dimensions that define things we are unlikely to encounter or even ones that could not possibly exist. In this sense, paradigms are not as efficient as some other kinds of arrangements. Consequently, they are rarely used for very complex typologies intended for highly diverse collections of data.

3.2.2 Taxonomic Classification

Taxonomic classification is the kind of classification that Linnaeus used to categorize plants and animals and that botanists and zoologists still use to classify organisms. Taxonomies work by making a series of distinctions (sometimes dichotomies), resulting in major categories that are subdivided into smaller categories, sub-categories, and sub-sub-categories. They are hierarchical and assignment of any item to a particular class is like running through a program from the "top" of the hierarchy on down. At each "level," the classifier makes a distinction on a nominal, dichotomous or ordinal scale, such as selecting "lithic" or "not lithic" or choosing among "small," "medium" and "large."

For example, one possible taxonomy for lithics could begin at the top with a distinction between tools and waste products. For a tool, the next level down might involve a distinction between core tools and flake tools (see Chap. 11). For flake tools, we might further distinguish between flakes that have retouch on the dorsal side only, the ventral side only, and both dorsal and ventral sides (bifacial). For bifacial retouch, we might distinguish ones with one retouched edge, two retouched edges, and more than two retouched edges. We continue down the hierarchy until the artifact is assigned to its proper category.

Among the properties of taxonomies, one of the most important is that they are hierarchical, with weighted criteria for assignment to classes. That is, criteria near the "top" of the hierarchy have more influence over the class to which an item will be assigned than ones near the "bottom." Another important one is that taxonomies are almost never symmetrical; all the "branches" are independent of one another and need not employ the same or even similar criteria for making finer distinctions (Fig. 3.2). For many of the things we would like to classify, this makes a lot of sense: why should we expect the criteria for subdividing lithic waste into smaller categories be the same as those for subdividing tools into tool types? How could they be the same? Another important thing is that taxonomies need not have any "empty" classes. We only need to employ distinctions that create classes for things we reasonably expect to encounter, at least rarely, and can omit ones that define classes for things that we are confident do not exist. In this sense, taxonomies are more efficient than paradigms (Table 3.3).

3.3 Grouping

Grouping offers more flexible arrangements than classification and is often useful either for exploratory work that will later lead to rules for a formal classification, or as a practical alternative that replaces classification altogether (Adams and Adams 1991).

Dunnell (1971) posits two kinds of grouping methods, or "non-classificatory arrangements," which he calls "numerical taxonomy" and "statistical clustering." Superficially, the former looks a lot like taxonomy but, in fact, it involves quite different procedures. The latter even begins with a paradigmatic classification. However, the variety of grouping methods that archaeologists, among others, use is actually much greater than Dunnell's dichotomy suggests, so here I propose a modified classification of grouping methods (Fig. 3.1).

Some of these methods are based on imposing boundaries, others involve attempts to reduce the number of dimensions of variability, and still others minimize the "distances" between group members.

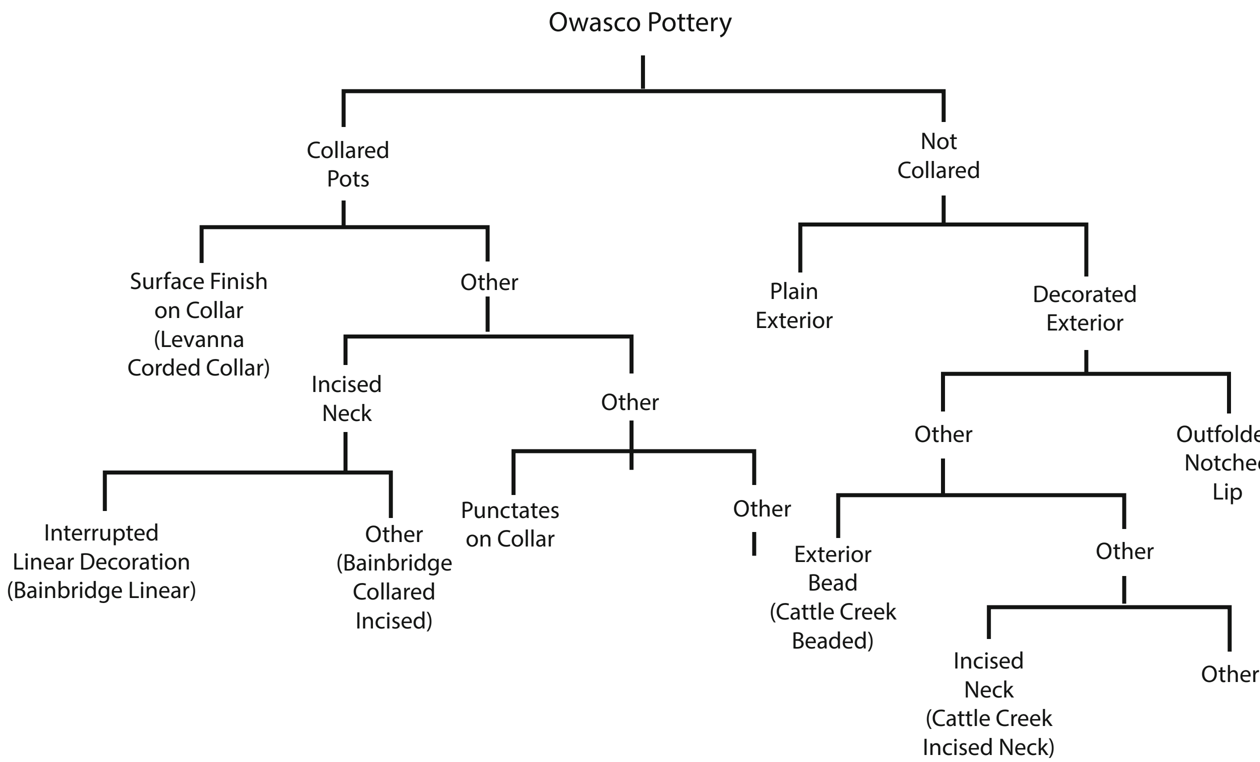


Fig. 3.2 A taxonomic classification of Owasco pottery from New York State (After Whallon 1972: 17). Note that the distinctions made in the left side of the "tree" are quite different from those on the right

Table 3.3 Key differences between paradigmatic and taxonomic classifications

Paradigms	Taxonomies
Dimensional	Hierarchical
Classes defined by intersection of dimensions	Classes defined by subdivision or distinction
Unweighted attributes	Weighted attributes
Symmetrical	Asymmetrical
May be inefficient	Highly efficient
Some classes may never have members	Classes not expected to have members can be eliminated
Unambiguous assignment to a class	Ambiguity can be avoided by strictly following a program from most inclusive to least inclusive classes

The grouping methods summarized below are common ones, most with a long history in archaeology. For a summary of more recent methods, many of which have yet to see as much application in archaeology as they probably should, see Baxter (2006).

3.3.1 Bounded Grouping

As already noted, archaeologists frequently and sometimes subconsciously group things by some boundary, typically a contextual (or extrinsic) boundary in space or time. **Extrinsic attributes** are ones that lie outside the objects being grouped or classified, such as their date of manufacture or the location where they were found. Intrinsic attributes are ones inherent in those objects, such as their length, chemical composition, or color. Bounded grouping is the backbone of just about any archaeological recording system, so much so that we tend to take it for granted. It also happens that, consciously or not, most archaeologists, especially when they are working in uncharted territory, begin a new artifact typology by grouping artifacts by context as the first step in another kind of grouping that involves looking for similarities and differences within and between groups.

I call this bounded grouping because membership in the group is usually described by the boundaries of a unit, typically the two-dimensional or three-dimensional space of all or some part of a layer, feature, excavation unit, site, or survey unit. The boundaries do not have to be spatial, however. We use chronological boundaries, for example, when we group artifacts of the seventeenth century.

One reason that this is not a classification is that classifications exist independent of space and time, while bounded groups depend on their space-time context. Another is that containment within the boundaries is a necessary but not sufficient condition for membership in the group. The boundaries of Layer 6, for example, might include many artifacts or "ecofacts" that we want to include in the group, but may also include things we do not want to include, such as the sediment matrix, non-cultural rocks, and so on.

As with any kind of grouping, the group description will change as new information accumulates. For example, continued excavation or analysis of materials from a particular context might identify more artifacts, thus enlarging the group and changing its statistical character.

3.3.2 Central-Tendency Grouping

This class of grouping methods places less emphasis on boundaries and more on central tendency or minimization of "distances" between group members and is typically based on the intrinsic attributes of group members. It includes a large range of sub-categories, from intuitive groupings based on an archaeologist's mental image of a "type" through more explicit statistical descriptions of group members to groupings based on distance measures or similarity coefficients.

A simple example involves keeping a reference collection of "ideal" examples of each category. Each object is then compared with these reference pieces and grouped with the one to which it seems most similar. This may sound a bit like classification, but really it is not, because the resulting units to do not exist independent of the objects and there are no rules that are either necessary or sufficient for assigning artifacts. Instead, assignment to a group is based on a sorter's judgment. This type of grouping is somewhat common for assigning pottery sherds to fabric types, in part because the detailed analysis of pottery fabrics is time-consuming and requires special training (see p. 197), while pottery collections are often very large.

Some more formal central-tendency grouping also has superficial resemblance to classification. Like the case of the reference collection, there could be "ideal" types to which each object is compared. A certain type of pottery, for example, may usually, but not always, have a red slip; it may tend to have an inverted neck, but with exceptions; it may typically have a rim diameter of about 15 cm, but the range might be as great as 12–20 cm. The ideal description of an assemblage of threshed wheat that was charred during parching could be that it has "lots" of spikelet forks, glume bases, rachis segments, prime grains and small weed seeds, but few culmn nodes or awn segments and no intact basal spikelets (see Fig. 16.1, Table 16.1). The "definitions" of these types are really descriptions, sometimes statistical descriptions, that summarize the characteristics of the most "typical" examples while recognizing that there is also variation.

Central-tendency grouping has other characteristics that plant it squarely in the realm of grouping methods. Even when they have structure that looks like a pre-arranged system, they begin with actual things, such as projectile points, pots or archaeological sites, that provide the initial examples to which later discoveries are compared. In fact, "type sites" that have provided the initial descriptions (and often names) of whole archaeological cultures are usually just the first such sites to be discovered, and later work often leads to shifts in the descriptions of these cultures such that the type site proves to be a poorer exemplar than other sites (e.g., Campbell 2007: 104–105). Consequently, the types are not abstract classes, and you would never have a type for which you would expect no members. Type descriptions of this kind are constantly refined as new examples are found and examined. For example, a type of pottery whose description originally included "usually red-slipped," might come to be "usually red- or black-slipped" as more black-slipped examples appear, or the type that included occasional black-slipped but mostly red-slipped ones might be split into two different groups once black-slipped sherds became more common.

Statistical Types and Type-Variety Typologies Many of the typologies that archaeologists have used for a long time are based on descriptions, sometimes statistical ones, of the modal characteristics of the type, with the expectation that actual examples vary around these modes. In some cases, we may keep a reference collection that exhibits the range of variation we would expect within each type and use it to match untyped-specimens. In other cases, we may publish descriptions and idealized pictures of each type (e.g., Garfinkel 1999; Goren 1992), or use a set of pictures that exhibits the range of variation in shape, to help us assign artifacts to morphological types. In much of Old World archaeology, for example, large numbers of published pottery drawings serve as "parallels" that archaeologists use to group newly excavated pottery with types known from other nearby sites (contra Adams and Adams 1991: 237). However, there are no explicit rules for using such illustrations for "typing" and usually the publications provide little or no statistical information on how much variability there is in each type.

Some cases of grouping things by the modes in a multimodal distribution may also belong to this type of grouping. For example, when there is a clearly multimodal distribution of site sizes in an archaeological landscape, we might use this as a basis for labelling sites close to each mode as "camps," "farms," "hamlets," "villages," and so on, although the modes and "troughs" in such distributions could sometimes be due to low sample size. This approach has worked better for more standardized artifacts and large sample sizes, such as standard weights and silver coins (Figs. 3.3 and 13.14), whose mass distributions often show clear peaks with patterned ratios to one another.

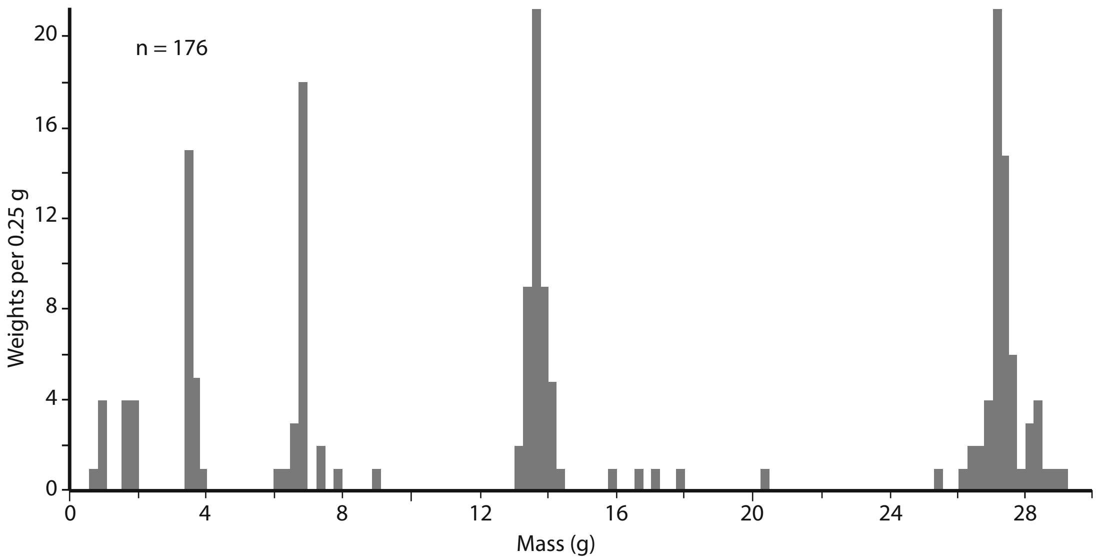


Fig. 3.3 Histogram of a multimodal distribution of Harrapan cube-shaped weights from the site of Mohenjo-Daro, Pakistan (Data from Mackay 1938: 607–611). Note that the distances between peaks tend to double, with peaks around 0.9 g, 1.7 g, 3.5 g, 6.8 g, 13.7 g, and 27.4 g, suggesting the system of values. Variation around these peaks, and scatters of intervening mass may be due to manufacturing error, dishonesty, chipping and wear, or error in the archaeologists' measurements

Polythetic Descriptions David Clarke (1968: 189–90) introduced to archaeological systematics the concept of polythetic definition to describe a kind of grouping that archaeologists had been doing for several decades, and biologists even longer (Sneath 1962). The key feature of polythetic descriptions is that they are based on a set of conditions or attributes, none of which is necessary or sufficient for attribution of any item to the group. Instead, we only expect each member of the group to share a large number of these attributes and each attribute to be shared by a large number of the group's members.

A classic early archaeological use of what we would now call polythetic description is V. Gordon Childe's definition of an archaeological "culture." He formally defined it as a complex or assemblage of regularly associated types that illustrate more than one aspect of human behavior, but recognized that no specific archaeological site belonging to a culture could be expected to exhibit all the important characteristics of that culture (Childe 1929: v–vi, 1956: 16, 33). For example, some kinds of artifacts, such as a particular kind of grinding stone, might occur at villages and farmsteads, but usually not in hunting camps or cemeteries, while others, like a particular kind of pottery vessel, might occur in villages and cemeteries but only rarely at farmsteads and never at hunting camps, even though all these sites belonged to the same culture. Childe viewed an archaeological culture as something exhibiting a constellation of attributes, only some of which would appear at each individual site, but each of which would occur at at least two sites belonging to that culture and be represented by more than one example. This comes very close to a polythetic definition.

Polythetic groupings are conceptually similar to the statistical types in that the description of each type, with respect to a number of attributes, is rather flexible. The main difference is that polythetic groupings use attributes on a nominal scale, allowing us to "score" each attribute as either present or absent in each site or feature or artifact. Members of the same group are similar to one another in the sense that they are identical (on some nominal scales) with respect to some large number of attributes, while differing in others. Polythetic grouping involves a clustering of nominal attributes, which distinguishes it from statistical grouping that clusters values measured on at least an ordinal scale.

One of the problems with polythetic description is an exaggerated version of one common to most grouping methods. Because the criteria for membership in a group, unlike the definitions in a classification, are flexible, there is no way to predict in what way any two members of a group may be similar or different. For example, items A, B, and C could constitute a group with attributes *abcdef*, *cdefgh*, and *efghij*, each sharing two-thirds of its attributes with at least one other member of the group, while the attributes *c*, *d*, *e*, *f*, *g*, and *h* occur in two-thirds of the group's members. Yet the ways in which A is similar to B are quite different from the ways in which B is similar to C, and A and C have only two attributes in common.

Despite this problem, some archaeologists would argue that polythetic descriptions come very close to the kinds of type "definitions" that archaeologists routinely use, whether consciously or not (Adams and Adams 1991: 226; Williams et al. 1973).

Distance Methods The class of grouping methods commonly called "clustering" and which Sokal and Sneath (1963) call "numerical taxonomy," has as its distinguishing feature the grouping of items by "distance" or dissimilarity between items in a multi-dimensional space. Central tendency for each group is achieved by finding a solution that minimizes the "distances" between pairs of group members. These are mathematical attempts to capture the kind of within-group similarity found in the intuitive and statistical variants of the statistical or type-variety methods and in polythetic descriptions. Common sub-classes of distance-based clustering include hierarchical clustering, optimal partitioning (including the k-means technique), density seeking, and multidimensional scaling (Aldenderfer and Blashfield 1984; Baxter 1994, 2003: 90–104; Drennan 2010: 309–320; Everitt 1974; Shennan 1988: 222–260; Sokal and Sneath 1963). Hierarchical clustering in a very simple form is a good way to illustrate most of the main principles of distance methods.

The essence of hierarchical clustering is to compare a set of objects with respect to a large number of attributes, and to group together those that are most similar to one another, while putting very dissimilar ones into different groups. This is based on "distances," which, in the simplest case, are just the proportion of disagreements when we compare pairs of artifacts with respect to an attribute list. The product of hierarchical clustering is a tree-like diagram that has superficial resemblance to a taxonomic classification, but it is actually quite different, since the "branches" are based on proportions of dissimilar traits or other measures of distance, not on defined distinctions.

Example

The following illustrates how a very simple example of hierarchical clustering works (see Shennan 1988: 219–240; Baxter 2003: 92–96) but is not intended as a recipe for how to carry out this kind of grouping, which would always involve software included in most standard statistical packages.

Often, hierarchical clustering begins with a long list of attributes measured on a dichotomous scale, such as "present/absent," "1/0" or "Y/N" (Table 3.4). We (or the software) compare items by counting the number of agreements (YY, NN) or disagreements (YN, NY) to yield a coefficient of similarity or dissimilarity for each pair of items, just the proportion of matches, or mismatches, among the total number of attributes.[2] Any pair of artifacts that agree on 75% of the attributes and disagree on the other 25% would have a similarity coefficient of 0.75 (or a *distance* of 0.25). In other words, if there were 100 attributes, they share 75 of them.

Once we have the coefficients, the goal is to group together items with high similarity (low distance) and separate ones with low similarity. However, when there are more than two items, how do we accomplish this if a third item shares many attributes with one of the items already in the group, but not the other?

The simplest, although not very good, method is called single-link clustering. This is an agglomerative, hierarchical method that serves to demonstrate how the coefficients can result in groups. We begin by searching a matrix of similarity or distance measures to find the pair that has the highest level of similarity (or lowest distance). In Table 3.5, this would be a similarity coefficient of 0.95 for the pair 5, 6. We then group artifacts 5 and 6 together at a level of 0.95, as in Fig. 3.4a, and search for the next-highest values and find that pair 1, 2 has a similarity of 0.9, so we add those to the graph as in

Table 3.4 Example of a matrix to record the presence (Y) or absence (N) of various attributes on six pottery sherds

	Artifact number					
Attributes	1	2	3	4	5	6
Collar	Y	N	N	Y	N	Y
Punctate collar	Y	N	N	N	N	Y
Incised neck	N	Y	Y	Y	N	N
Interrupted lines	N	Y	N	N	Y	Y
Exterior punctate	N	Y	Y	N	Y	N
Exterior cord markings	Y	N	Y	Y	Y	Y
Notched lip	N	Y	Y	N	N	N
Punctate lip	Y	N	N	Y	N	Y
Exterior herringbone	Y	Y	Y	Y	Y	Y
Exterior horizontal lines	N	N	Y	N	N	N
Exterior oblique lines	N	Y	N	Y	Y	N
Exterior vertical lines	N	N	Y	N	N	N
Castellation	Y	N	N	Y	N	Y
Sharp shoulder carination	N	N	N	N	N	Y
Lines on lip interior	Y	N	N	Y	N	N
Rouletted lip	N	Y	Y	N	Y	Y
Wavy rim	Y	Y	Y	N	Y	Y
Effigy on collar	Y	N	N	N	N	N
Ridge below collar	N	N	N	N	N	Y
Punctate row below collar	Y	N	N	N	N	N

Note that a "castellation" is a raised portion of a collar. Can you detect what is wrong with this matrix?

(continued)

[2] Note that some kinds of clustering use the "Jaccard coefficient," which ignores NN matches. This can be advantageous in many archaeological applications when the absence of an attribute is probably due to small sample size.

Table 3.5 Example of a similarity matrix for seven artifacts (After Orton 1980: 48–49)

	Artifact Number						
	1	2	3	4	5	6	7
1	1.0	.90	.80	.65	.40	.40	.30
2		1.0	.85	.60	.40	.30	.30
3			1.0	.50	.25	.25	.20
4				1.0	.70	.60	.55
5					1.0	**.95**	.80
6						1.0	.80
7							1.0

The bolded rectangle marks the highest similarity score

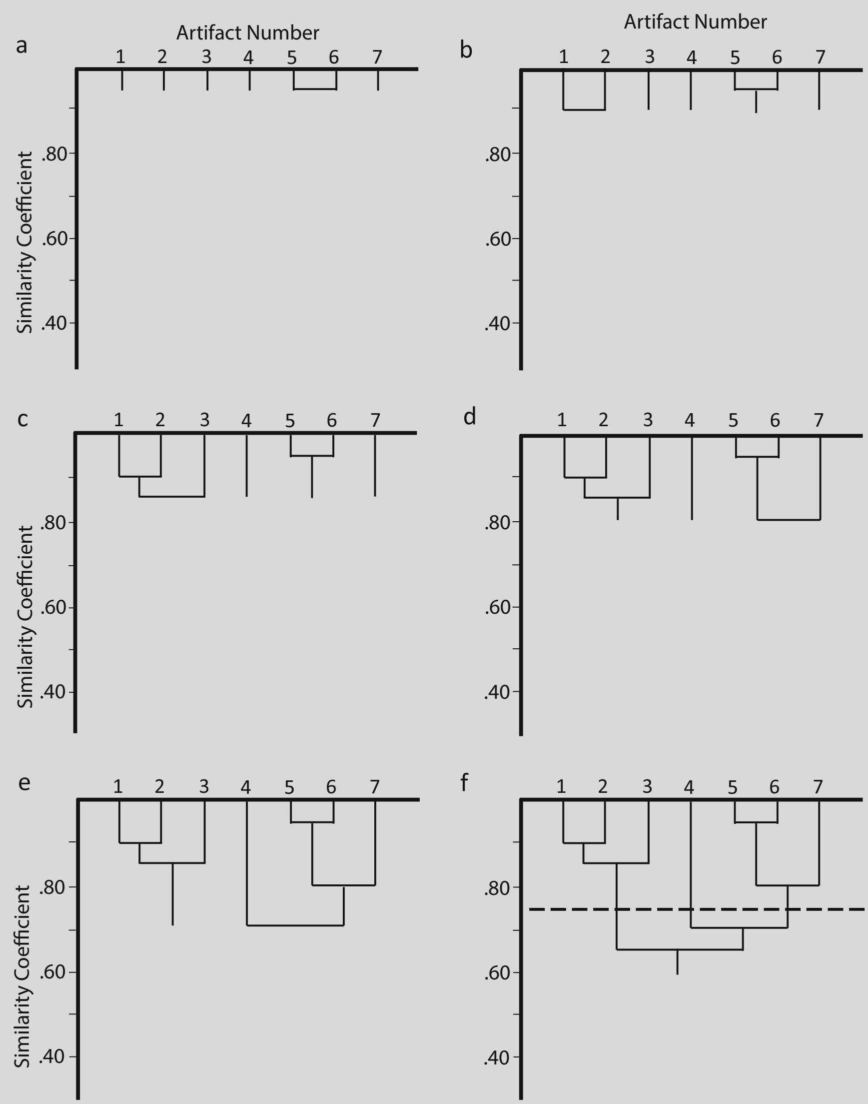


Fig. 3.4 Steps in single-link clustering using the similarity matrix in Table 3.4. (After Orton 1980: 48–49): (**a**) artifacts 5 and 6 are grouped at 95% similarity, (**b**) artifacts 1 and 2 are grouped at 90% similarity, (**c**) artifact 3 joins the group with artifacts 1 and 2 because it has 85% similarity with artifact 2, (**d**) artifact 7 joins the group with artifacts 5 and 6 because it has 80% similarity to both 5 and 6, (**e**) artifact 4 joins the same group because it has 70% similarity to artifact 5, and (**f**) setting a "cut-off" of about 75% defines three groups with the sets (1, 2, 3), (4), and (5, 6, 7)

(continued)

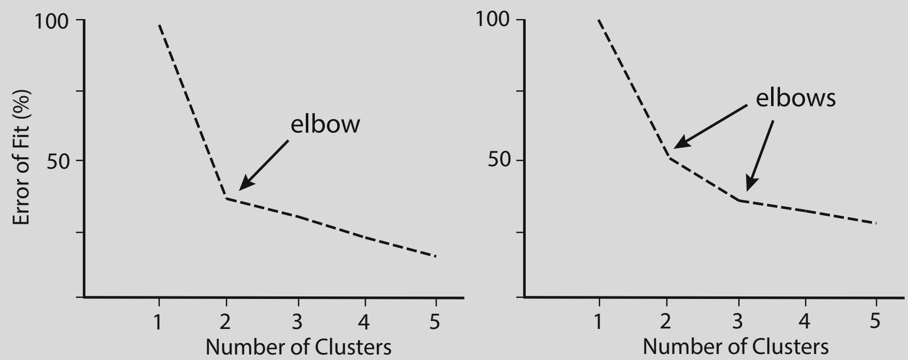


Fig. 3.5 Plotting the error of fit in the *k*-means technique to find the best number of clusters. In (**a**), the elbow occurs at two clusters ($k = 2$), while in (**b**), there are two elbows and, arguably, the best number of elbows could be three. (After Orton 1980: 53)

Fig. 3.4b. The next-highest is 0.85 for the pair 2, 3. Should we add artifact 3 to the group already containing sherds 1 and 2? In single-link clustering, we do, as it only requires that level of similarity to at least one member already in the group, and we ignore the fact that artifact 3's similarity to artifact 1 is only 0.8. We continue this process until all the artifacts are linked up at some level of similarity, resulting in a tree-like graph, or dendrogram, that represents the hierarchy of similarity or dissimilarity. Finally, we select a "cut-off" for the minimum similarity we will use to define the groups. The horizontal, dashed line in Fig. 3.5f represents one such choice, indicating that there are three groups: (1, 2, 3), (4), and (5, 6, 7). While we can select this cut-off subjectively, there are also methods for deciding how many groups there should be and where to place the cut-off.

As with polythetic descriptions, single-link clustering yields groups whose members' shared characteristics and dissimilarities are unpredictable. Just because a particular item is sufficiently similar to at least one member of an existing group to justify adding it to that group does not mean that it is especially similar to other members of the group. To return to the hypothetical artifacts A, B, and C, with sets of attributes *abcdef*, *cdefgh*, and *efghij*, single-link clustering could add a fourth member, D, with attributes *ghijkl*, just because it shares two-thirds of its attributes with C (similarity of 0.67). However, it would share only one-third of attributes with B and no attributes at all with A. This problem is sometimes called "chaining" because you can have a series of items linked by similarity, but items at either end of the chain that are not similar at all. This is why single-link clustering is a poor choice.

Agglomerative methods such as double-link, total-link, or average-link clustering, "Ward's method," or some recent alternatives are therefore preferable. These require that each new member of a group has at least some minimum level of similarity with at least two, or all of the existing members of a group, or that membership is based on the average of all the similarity coefficients of all possible pairs in the group. Ward's method agglomerates in such a way as to minimize the increase in intra-group variability (or "sum of squares") when items are added to the group. This makes it less likely that an added item will have practically nothing in common with some other members of the group, but still does not guarantee that any two members will be greatly similar.

A fairly common, and quite serious, error is illustrated by the example data in Table 3.4. For any distance method to work well, the attributes should be independent of one another. Sloppy thinking in the selection of attributes causes some of them to have greater weight than others, just by accident. In this example, note that one of the attributes is the presence of a "collar" on the sherds (a thickening of the rim that is a frequent attribute of Iroquoian sherds). However, several other attributes on the list are actually attributes of the collar. Consequently, any sherd that scores "N" for collar must also score "N" for such attributes as "punctate collar," "castellation" (a raised feature on the collar), "effigy on collar," and "punctate row below collar." Consequently, the mere presence of a collar has a huge influence on the resulting groups. That should never happen just by accident.

Another somewhat disconcerting characteristic of numerical clustering is that, while each member of a group has a certain level of similarity with other members of its group, the ways in which it is similar to one member can be completely different from the ways it is similar to another, as we have already seen in the case of artifacts A, B, C and D. There is no way to predict the ways any two members will be similar, although Ward's method at least ensures that there will be more points of similarity. Consequently, the results of numerical clustering are really polythetic sets, even though we use different methods to create the groups.

The most serious problem with average-link and similar hierarchical agglomerative methods is that a small change in one of the coefficients could result in a substantial change to the whole dendrogram, not just one or two of its branches (Jardine et al. 1967). If there is more than trivial uncertainty in the calculation of the coefficients (e.g., from missing data or the selection of attributes), this could be a serious defect, and is one reason to prefer Bayesian methods (below).

Should you use one of these methods, then, you should define the attributes you will use very carefully and try to avoid redundancy or interdependence of attributes (Read 1982). You should also try to avoid using attributes that almost never occur, leading to a Y/N score of "N" in almost every case, unless you use a coefficient that does not take N-N matches into account (Jaccard coefficient). There can be many other cases of highly correlated attributes, a problem that archaeometrists trying to cluster artifacts by their chemical compositions have recognized for a long time.

The problems just mentioned have led many archaeologists to favor other methods that are not only based on pairs of items. Hodson (1970) introduced the ***k*-means technique** or "locational clustering" (Kintigh 1990) to archaeology. This method partitions the items into a specified number (k) of clusters in such a way as to minimize the squared distance (dissimilarity) between each item and the centre of its group in "space." By this method, two items that might be in the same group when there are three groups ($k = 3$) might be in different groups when there are four ($k = 4$). You must decide what level of k seems reasonable, or repeat the method with different values of k. You can then plot a graph that shows how the relative error of fit (the average squared distance of the items from the group centers as a percentage of their average squared distances from the center of all the items) is affected by increasing the number of groups, k (Fig. 3.5). The "best" choice of k is usually where we find a sharp bend or "elbow" in the graph, indicating that there has been a marked improvement in fit over lower values of k, but that increasing k further would result in only modest improvement.

Multidimensional Scaling (MDS) is another alternative among the distance-based methods. It more literally takes advantage of the characterization of dissimilarities between artifacts as distances in a multidimensional space. Because they are multidimensional, we cannot accurately portray these distances by points on a map-like graph, because those are two-dimensional. If, however, we are willing to accept some distortion, much as we routinely distort Earth's spherical geography to fit onto two-dimensional maps, we can illustrate these distances at least crudely. For example, we might allow a distortion so long as it correctly represents the rank-order of the distances (Baxter 2003: 85–86; Orton 1980: 55). Sometimes we must distort the map still further so that the points on it will fit (Fig. 3.6). As with numerical taxonomy, many statistical packages will do multidimensional scaling.

Other Numerical Clustering Methods Alternatives to the methods outlined above include ones based on partitioning of items or density searching rather than hierarchical aggregation or division but share the concept that we can consider the items as points in a multidimensional space. In all these cases, the goal is to minimize distances between these points inside the groups, while having larger distances between members of different groups (Aldenderfer and Blashfield 1984; Baxter 1994).

While distance techniques are useful, they have also disappointed archaeologists who had hoped that they would provide an "objective" way to do archaeological systematics.

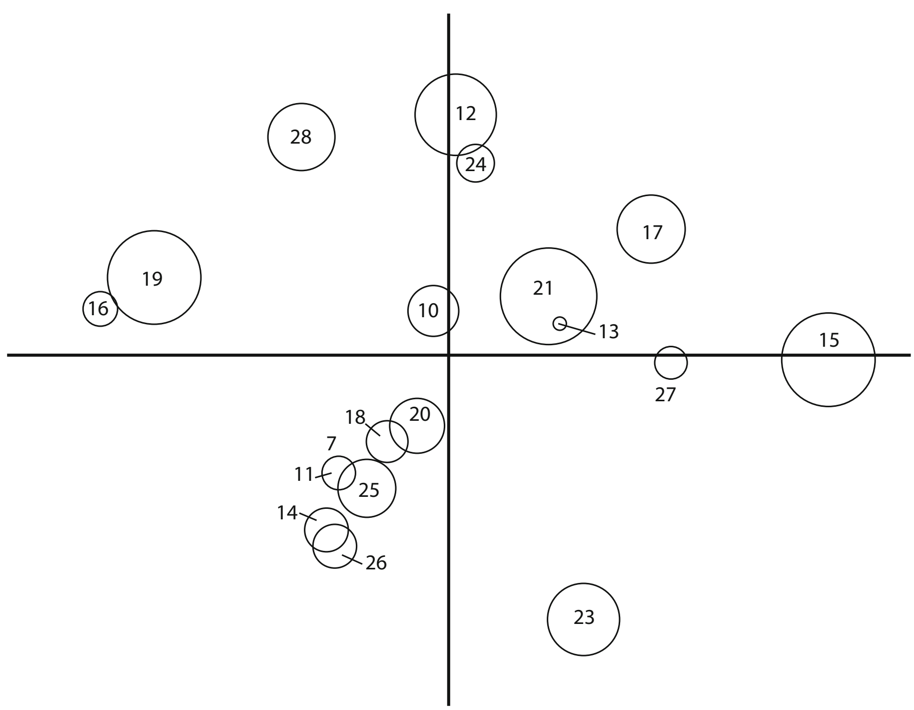


Fig. 3.6 Two-dimensional MDS "map" to illustrate the dissimilarities (distances) among the locational attributes of Late Iron Age sites in coastal Zululand, South Africa. The diameters of circles represent a third dimension, the numbers represent different site attributes, and the stress (distortion) is 6%. (After Hall 1982: 140)

What you will find is that the results can be very sensitive to your decisions about what attributes to measure, how many to use, and how to measure them, errors in measurement, how many groups there should be, and additions of artifacts that were not originally included. One way to deal with this problem is to re-do your analysis several times with slightly different assumptions, or longer or shorter lists of attributes, to see how well your groups hold up. Going with the first groupings that result and not doing any checks may make your results suspect.

3.3.3 Attribute Association

The other major class of grouping methods that Dunnell (1971: 95–98) recognizes is one he calls "statistical clustering" but I prefer to call it "attribute association" to avoid confusion with the distance methods just described.

In attribute association, identification of groups depends on statistical associations between attributes. This is an extension of archaeologists' intuition that an artifact "type" involves some recurring combination of attributes (something that also happens with paradigmatic classifications and polythetic sets). In that case, we might expect the attributes to be associated or correlated.

For example, one artifact type in Classical archaeology is the amphora, which consistently has an elongated form that tapers to a nearly pointed or knob-like base, a neck much narrower than the body, and two handles placed vertically on either side of the neck. Although other kinds of pots had one or two of these attributes, the strong association of these four attributes in amphoras allows us to predict un-observed attributes from some of the others. If we were able to depict a four-dimensional "cube" with "sides" partitioned by the presence and absence of each of these attributes, we would find that amphoras all cluster in one space in the "cube" while other kinds of pots (e.g., some with pointed bases but no neck or only one handle) would cluster elsewhere or not make strong clusters at all.

An informal kind of attribute association has had long use for grouping artifacts or raw materials by their elemental or isotopic compositions. For example, archaeometrists make x/y scatterplots with either the concentrations of elements, or ratios between two elements or isotopes, on the two axes. The resulting dots might make a meaningless scatter, or follow a diagonal line to indicate a correlation, but often they exhibit several fairly distinct clusters that indicate artifacts or raw materials that are similar in composition, at least with respect to the elements or isotopes represented in the graph (Fig. 3.7). By analyzing both artifacts and raw materials from known sources, this approach can be used to

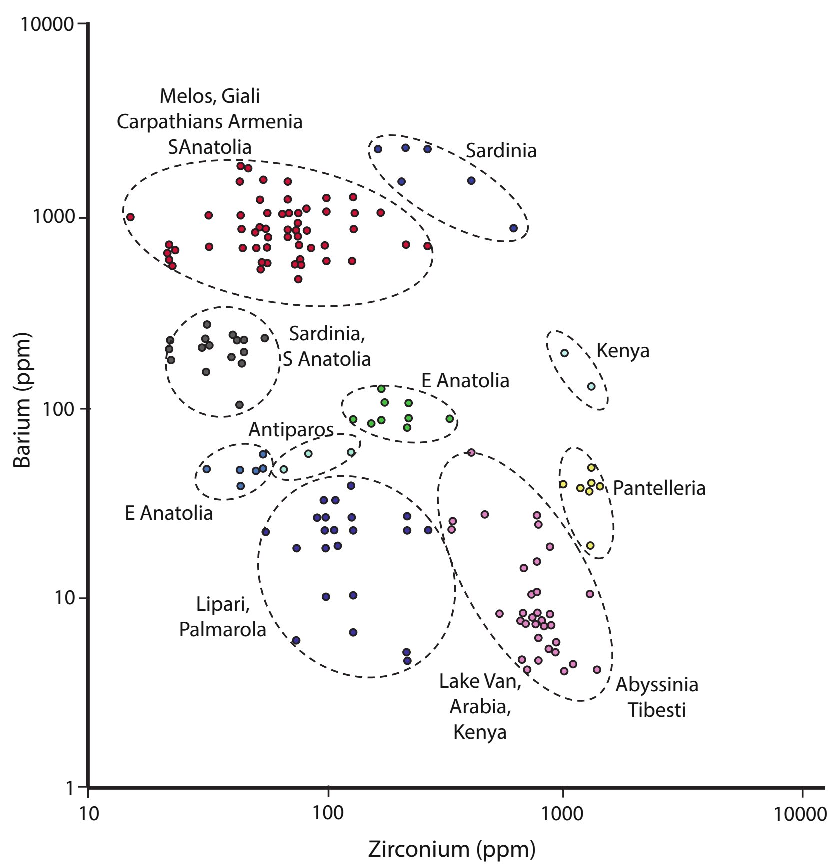


Fig. 3.7 Trace-element concentrations (parts per million) of barium and zirconium in Old World obsidians of known sources. (After Cann et al. 1969: 157). Note the logarithmic scales

"fingerprint" sources and trace artifacts to those sources (artifacts falling within the same group as samples from a known source). In other words, we can predict the value of an **extrinsic** attribute (source of an artifact's material) from the artifact's **intrinsic** attributes (e.g., chemical composition). This approach has been successful at sourcing things like obsidian (trace elemental composition) and silver (lead-isotope composition), as discussed further in Chaps. 11 and 13. The groups associated with each source in this case are sometimes easy to see in these simple graphs of attribute associations. However, you will note that the short between-point distances within groups and longer distances between groups mean that we could also have found these groupings with one of the distance methods. Consequently, archaeometrists often use such graphs in an exploratory way but, for further analysis, use either one of the hierarchical distance-based methods already mentioned or one of the following methods.

The methods that follow do not formally identify groups in the way that the distance-based clustering methods do but can be used to derive groups or group membership when, as in the simple examples just mentioned, recurring associations among attributes are important contributors to the constitution of groups.

Factor Analysis This is a method that works well on data that include considerable correlation or covariation among attributes (Lawley and Maxwell 1971). Factor analysis expresses the data with fewer dimensions by reducing them to a few composite dimensions, or "factors," that are something like summaries of several co-varying or correlated attributes (Shennan 1988: 271–280). When dimensions are highly correlated (either positively or negatively), it implies that they may just be aspects of some other dimension. To give a very simple example, where length, width and mass are correlated, they are really just aspects of size, so factor analysis would reduce these separate dimensions to a single size dimension. In other cases, the exact nature of the factor that entails the correlated dimensions may be much less obvious or include aspects that are negatively correlated with others. For example, the chemical compositions of pottery will show some positive correlations between elements because those elements combine in constant ratios in one of the minerals that the pottery contains. By contrast, when the data consist of proportions of something like pollen or animal bones, any increase in one proportion necessarily entails decreases in other proportions, causing some negative correlations. For example, the contribution of deer might be negatively correlated with that of fish, indicating that they are both aspects of a dimension concerning a marine or terrestrial diet.

Principle Components Analysis PCA also makes use of the associations between attributes that show inter-correlation or covariance, making it somewhat like Factor Analysis. Its principle is to transform the data linearly and orthogonally (at right angles) into a set of new dimensions, or "components," that are uncorrelated and in such a way as to maximize the variance of the first component, while the second component has the second-highest variance possible while being uncorrelated with the first component, and so on. This allows the first few components to account for most of the variance in the data, and we can plot their values on scatterplots. Because the axes of the graph are now oriented along the axes of maximum variance, we sometimes get better separation of any "natural" groupings in the data (Baxter 1994: 48–99; Drennan 2010: 299–307; Shennan 1988: 245–270), but there is no guarantee that such clustering will appear. Because archaeologists use PCA in seriation more often than for grouping, we will return to this method in Chap. 18.

Correspondence Analysis CA is similar to PCA except that it is used with counts on nominal or dichotomous scales, rather than interval measurements. In archaeology, it has most often been used for exploratory analysis of tables of artifact counts by context (Baxter 1994: 100–107, 2003: 136–146; Shennan 1988: 283–286). It helps users identify associations between the rows and columns of the table. Plots that result from CA often have a characteristic curvature when the data have a serial order but may form clusters otherwise (Fig. 3.8). Like PCA, it is often useful in seriation (Chap. 18) but can sometimes be used to identify groups.

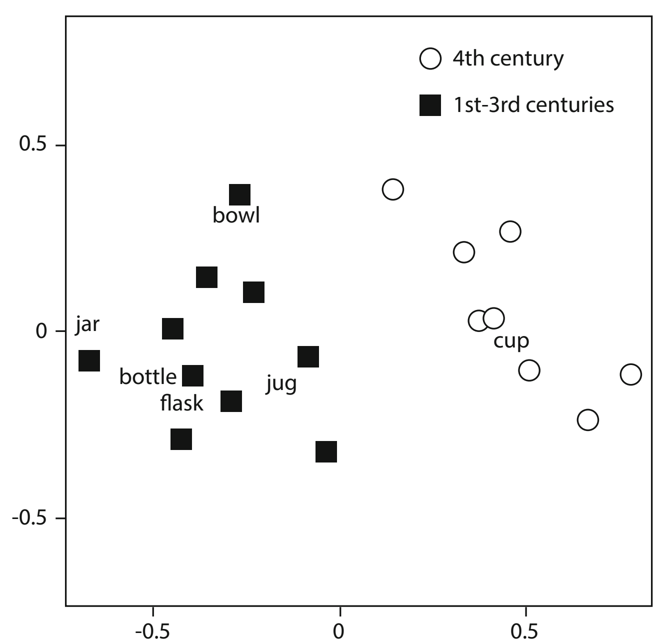


Fig. 3.8 Correspondence Analysis (CA) of assemblages of glass vessels (After Baxter 2003: 138). The data fall into two clusters of earlier contexts (first to fourth centuries AD), dominated by bottles, flasks and jugs, and later contexts, dominated by cups

Configurationist Typology An old version of attribute association involves looking for patterns in data at the nominal or dichotomous scale in the hope of discovering "natural" types. Spaulding (1953: 305–306) argues that "a pronounced association of two attributes is the minimum requirement" to demonstrate that a type exists, and a type is "a group of artifacts exhibiting a consistent assemblage of attributes whose combined properties give a characteristic pattern."

This method begins with a paradigmatic classification of artifacts from a single assemblage, then tests the distribution among classes for statistically significant associations between attributes. Adherents of the Spaulding approach argued that statistical tests like chi-square would determine whether "valid types" exist in the classification (Watson et al. 1971: 126–132). A type in this sense only exists if there is a non-random relationship between the attributes, meaning that, for any artifact, we can predict the value of one attribute as long as we know the value of another (Table 3.6a). Notably, if we were to reduce the ratio-scale data in Fig. 3.9 to dichotomous scales partitioned by the dashed lines, the method would "verify" a typology of long, wide and short, narrow projectles in *a* and *b*, but would fail to identify the two clear clusters in *c*, which have different ratios of length to width. As Hodson (1982) points out, this approach would not verify even very strongly confirmed typologies, such as commonly used classifications of social roles by gender and age, and the Spaulding method is now obsolete.

However, the attribute association method is more defensible when at least one of the attributes is *extrinsic*, rather than intrinsic to the artifacts themselves. As archaeologists have long recognized, non-random associations between some intrinsic attributes, such as projectile form or pottery decoration, and extrinsic contextual ones, such as stratigraphic position, functional context, or spatial location, can identify types that are meaningful, at least with reference to a particular archaeological research question (Table 3.6b, c; e.g., Gilboa et al. 2004: 681, 687).

3.4 Misclassification and Uncertainty in Systematics

Although classes in classifications have rules that are supposed to make the classification process unambiguous, this does not mean that items are always classified without error. For example, we routinely classify things like pottery slip or flint raw material by color, with nominal categories like "red" or "brown" or ordinal categories using a Munsell chart or its digital equivalent (see p. 302). Yet different observers frequently disagree on these assignments, sometimes hedging by using additional classes like "reddish brown" but also because one of the observers simply makes a mistake, or two observers see things somewhat differently. The mistake would typically involve failing to follow the rules that are part of the definition (with color, because we use that classification so routinely in daily life, we usually do not explicitly think about what those rules are).

A thorough Quality Assurance protocol for an archaeological laboratory should deal with misclassification as an aspect of measurement error. A very simple measurement of this error is called the **error rate**; it is just the proportion of incorrect classifications as determined by an expert classifier, an approach to error that, unfortunately, depends on the expert being infallible and has some other shortcomings (see Chap. 1).

Table 3.6 Attribute association for configurationist typology (a) disregarding context (intrinsic attributes only, after Watson et al. 1971: 128), and with stratigraphic context (extrinsic attribute) in (b) and (c)

a		Dimension 1 (Temper)	
		Shell-tempered	Not Shell-tempered
Dimension 2 (Painting)	Black-on-Red	26	25
	Not Black-on-Red	22	27
b	**Stratum I**		
		Dimension 1 (Temper)	
		Shell-tempered	Not Shell-tempered
Dimension 2 (Painting)	Black-on-Red	25	3
	Not Black-on-Red	11	20
c	**Stratum II**		
		Dimension 1 (Temper)	
		Shell-tempered	Not Shell-tempered
Dimension 2 (Painting)	Black-on-Red	1	22
	Not Black-on-Red	11	7

Note how the data from (a), when distinguished by stratigraphic context, now show clear patterning, with high counts on one diagonal and low counts on the other

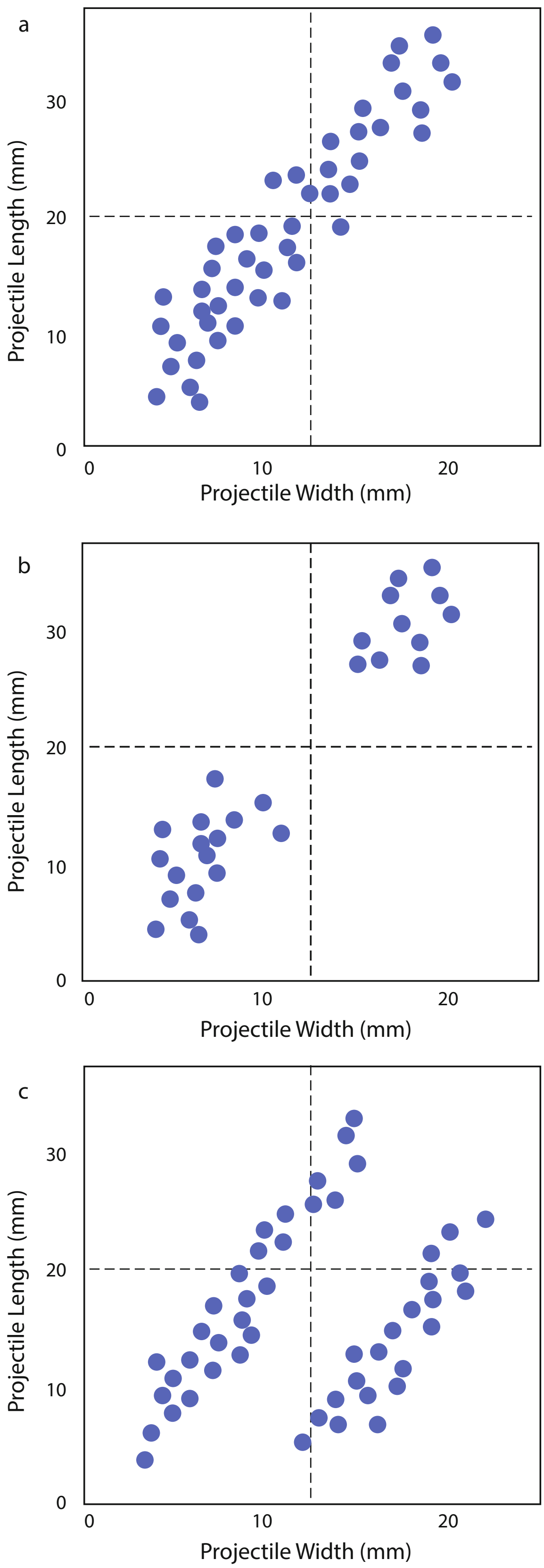


Fig. 3.9 Three different relationships between the length and width of fictitious projectile points, with dashed lines dividing the field into four quadrants analogous to the cells in a paradigm with two dichotomous dimensions (2 × 2). In (**a**), there is a correlation but no distinct clusters, in (**b**), there may be two distinct clusters or just a correlation with missing data, and in (**c**), there are two distinct clusters with different length/width ratios

3.4.1 Fuzzy Classification and Fuzzy Sets

Fuzzy classification and fuzzy grouping deserve more attention in archaeology as they may describe some aspects of real archaeological typologies (Adams and Adams 1991: 72; Baxter 2009; Coppi et al. 2006; Niccolucci and Hermon 2015). Its basic approach, rather than assuming that items can be assigned to classes or groups without error, is to assign probabilities that the item belongs to one or more classes (or estimates of the degree to which the item fits each class) or to spread membership among two or more groups. If a classification is certain, its probability is 1.0. In other cases, however, a pottery sherd might have a 0.6 probability of belonging to the "bowl" class but a 0.2 probability of belonging to the "jar" class and 0.2 probability of being something else.

Note that this contradicts the requirement in "normal" or monotethic classifications that assignment of an item to a class be unambiguous but is more realistic in that it anticipates that some assignments are uncertain. In archaeological contexts, this can be due to incompleteness of an artifact, difficulty in measuring some of its attributes, strong similarity between meaningful taxa, such as sheep and goats in zooarchaeological analysis, or even problems with the classification itself.

However, fuzzy classification is not polythetic in the sense of the grouping methods discussed in this chapter. That is because the classes still have definitions, rather than descriptions, and the fuzziness is simply a function of our uncertainty in assigning items to classes (see Chap. 1).

3.4.2 Fuzzy Grouping and Bayesian Grouping as Responses to Uncertainty

Just as we can have fuzzy classification, it is also possible to do fuzzy clustering (Döring et al. 2006), which uses statistical methods similar to some of those above, especially k-means clustering, but assigns each item degrees of membership to each of the clusters (usually summing to 1.0 to resemble probabilities). A key feature is that analysts must decide the value of a "fuzzifier" or "fuzzification" factor; where this factor is 1.0, a result identical to that from traditional clustering obtains, but values much over 2.0 produce much fuzzier results, with "softer" boundaries between clusters. To date, this has found few applications in archaeology, and some of these concluded with a "hardening" of results by assigning items to the groups for which they had the highest membership values. As Baxter (2009: 1039) notes, doing this "possibly misses the point of doing fuzzy clustering."

The Bayesian approach (see pp. 139–140) is another way to deal with the potential uncertainty in clustering items into groups (Buck et al. 1996: 312–325). This allows us to determine the probabilities of different numbers of clusters (*k*, as in k-means) and the most likely assignments of items to each cluster. As an example, Dellaportas (1998) uses Bayesian modelling with prior information and Markov-Chain Monte Carlo (MCMC) methods, which involve random sampling of a probability distribution, to find posterior probability densities for clustering Neolithic stone tools into two groups, axes and adzes.

3.5 "Automatic" Classifications and Artificial Intelligence

For many decades, archaeologists have dreamed of objective typologies that were not vulnerable to archaeologists' preconceptions. The review above should convince you that truly objective ones are illusory, yet some advances in computer pattern recognition may conceivably allow us to identify groups or assign items to classes more consistently, with less error.

Although most of the research on this topic is non-archaeological, some recent advances in this area include classification of pollen grains (Gonçalves et al. 2016) and even automated detection of the writers of ancient Greek inscriptions (Panagopoulos et al. 2009).

3.6 Practical Considerations

Given the wide array of options, it is important to consider practical aspects that might lead us to choose one or another form of systematics in our research design.

Among these is ease of use. How easily or consistently can we measure the attributes needed to assign things to classes or describe groups? Some attributes that might be very useful may only rarely be preserved on fragmentary specimens, and it is usually more helpful to use ones that are regularly observable on almost every piece. Meanwhile, a potentially useful attribute that only very highly trained observers can consistently measure, or that requires expensive equipment or very time-consuming observation might best be avoided unless it is critical to your research design (Adams and Adams 1991: 237).

Another is redundancy. For example, there is no point in conducting expensive or time-consuming measurements if we can achieve the same groupings with fewer, cheaper and easier attributes (Adams and Adams 1991: 236). Only use as many attributes as you really need to accomplish your research goals.

In a similar vein, you do not need to elaborate your typology so much that there are thousands of types. Each type should have a purpose (Adams and Adams 1991: 242). In a classification, it is usually preferable to have classes to which we expect to assign many items, rather than having many "empty" classes. Initial exploration of data may require us to have more categories, but we can later eliminate ones that contribute little or nothing to accomplishing our research goals. Taxonomic classifications tend to be very amenable to combination of little-used categories. However, sometimes we also retain certain categories from convention or because we want to compare our data with that of other archaeologists who have used them, either recently or in the more distant past.

Another factor is distinctiveness. It is preferable to have types and attributes that allow us to be confident and precise about their meaning, with respect to the typology's purpose. For example, in a chronological typology we favor artifact types that only occurred for a short period over ones that continued with little change over many centuries. Archaeologists sometimes call the former types "highly diagnostic" or "type fossils."

It may also be preferable to have types based on attributes that are clear and reasonably reproducible from observer to observer rather than ones with "fuzzy" boundaries that are vulnerable to ambiguity, as long as their attributes are just as useful with respect to the typology's purpose. Hand (1997: 99, 109–115) describes the desirable characteristic of having easily distinguishable categories as **separability**. Separability is high whenever categories are perfectly separated, as happens when certain attributes are uniquely associated with one class in a classification. Grouping methods, because they are polythetic, or have overlapping sets of attributes, can have low separability, not to mention fuzzy clusters.

Some practical aspects concern how our approach to systematics meshes with our use of digital information systems. Most databases require us to have unambiguous and unchanging definitions for the entities they will document, and this requires us to have developed thoroughly either a classification system or a set of groups with very high separability. Having a database with fields whose definitions changed from one excavation season to the next, for example, would result in many problems of data interpretation. The usefulness of the database for making comparisons and detecting patterns depends on consistency (see Chap. 4).

A general principle is to keep it simple. Typologies should not be more complicated than they need to be to achieve research objectives (see Chap. 6).

A final practical consideration is the level of consistency and accuracy we expect from the use of a classification. This is a matter of quality.

3.7 Quality in Typologies

Although most archaeologists would like all the people associated with their project or laboratory to assign material to types in the same way, we must anticipate some level of error, decide how much is tolerable, and find ways to keep errors within this limit. We do not need to expend resources on reducing error below that limit, but we should report estimates of what the errors are (e.g., Adams and Adams 1991: 238).

As noted in the last chapter, one measure of error in (non-fuzzy) classifications is **error rate**, often expressed simply as the estimated proportion of incorrect assignments to a class. This approach has two problems, however.

First, this approach treats all errors as equally serious (Nance 1987: 258–267). For example, the error of misclassifying a Combed Beaker as a Corded Beaker is considered the same as classifying a Corded Beaker as a bowl. More sophisticated measures of misclassification treat some kinds of error as more serious than others. For example, misclassifying a Corded Beaker as "unknown" may be less serious than misclassifying it as a Combed Beaker, and measures of error might refer to the expected costs or risks of different kinds of misclassification, or may employ a "confusion matrix" for the classification rules (Hand 1997).

Second, in archaeological contexts, use of any error-rate approach usually makes it necessary to assume that some expert's classifications are without error and count as a misclassification any observation that deviates from the expert opinion. As this is not very realistic (even experts make errors from time to time), it does not account for all potential errors in classification, and most archaeologists have instead opted to evaluate error through some measure of reliability (see p. 11). However, use of a control chart that tracks difference between lab personnel and a supervising expert can help us explain sudden changes in error rate, for example, whether they are due to changes in either the expert or the personnel under that expert's supervision (Fig. 12.29).

An alternative that does not entail the assumptions in error rate is to do statistical tests for reliability by having multiple analysts classify the same artifacts or ecofacts.

Prentiss focuses on the validity[3] and reliability of a widely used typology for lithic debitage. However, Prentiss was explicitly "not interested in . . . inter-observer error" (Prentiss 1998: 638), but instead in the ability of the typology to classify debitage from the same technological process consistently to the same category. PCA served to assess the typology's ability to assign flakes correctly to either tool production or core reduction, using artifacts from experimental lithic reduction, and several varieties of the production modes (e.g., prepared core, see Chap. 11). The research design also accounts for several potential confounding variables in the reduction events. Controls included holding knapper and raw material constant and randomizing the assignments of cores and blanks and the timing of the reduction events. The conclusion is that the typology had high reliability, as flakes were assigned consistently, with little random error, but low validity, as the typology was not effective at distinguishing tool manufacture from core reduction (low concurrent validity).

Whittaker et al. (1998) discuss some general issues regarding typological reliability and then their test of consistency among 13 archaeologists of varying experience in their classifications of pottery from the northern Sinagua region of Arizona. To accomplish this, they use a method called consensus analysis (Weller and Mann 1997; Weller and Romney 1988), which examines patterns of agreement among the classifiers or, more precisely, the degree to which their assignments to classes indicate a shared use of a model for the classification, or whether there were apparently two or more models or just "turbulent" assignments, indicating no shared models at all. It produces a "key" sorting solution only when there is a high level of agreement among sorters. In their demonstration, they find that the classifiers adhered to a single model, but with considerable variation. However, the consensus model was a "lumping" one, with only a few broad classes; when sorters attempted to make finer distinctions, not surprisingly, there was less agreement.

Statistical grouping methods also require evaluation of the quality of their results. Some of the methods are very sensitive to outliers or noise in the data, so it is important to check whether outliers are having undue impact on results by seeing how much difference it makes when you omit them. Another problem is that some clustering methods tend to impose spherical clusters on the data, when the real clusters are elliptical or some other shape, and even when there is no patterning in the data at all (Baxter 2006). If there is reason to suspect that this could be a problem with your data, you should use a method that performs better at identifying non-spherical clusters.

3.8 Do Typologies Have Real Meaning?

As noted especially in connection with configurationist typology, archaeologists have long asked themselves whether the types we create have any real meaning in terms that members of a prehistoric culture would recognize. Sometimes, they have (often incorrectly) painted this in terms of an emic:etic distinction, "emic" being the term for the perspectives,

[3] Prentiss initially uses the term "validity" as a synonym for accuracy, or lack of bias or systematic error. However, we would expect a valid measurement also to be accurate, and Prentiss later makes use of a version of construct validity as well as, implicitly, concurrent validity.

systematics, and world-views of a cultural group that is the subject of study, and "etic" for the perspectives, systematics, and world-views of the researchers studying that group. All archaeological analyses are inescapably etic because we do not really access the thoughts or concepts of past people, only detect residues of their actions. There is certainly a temptation to interpret typological distinctions and typological change over time as consequences of social relationships, identity formation, political actions, and communities of practice. However, archaeologists who have explored such themes in ethnographic contexts have found that simplistic social-interaction models can be far from the mark (e.g., Hodder 1977), even if they do seem to hold up in some other contexts. Even in cases where the associations between archaeological types and ancient social categories are strong, our types are still modern and, in some sense, arbitrary categories.

Some archaeologists question the use of typologies at all. Typologies inevitably impose structure on the observations, homogenizing diversity and reducing difference by ignoring an infinity of characteristics, privileging a small set of attributes, and assigning individuals to classes or groups in ways that, some argue, "erase" their individuality and provide an unjustified impression of certainty (Boozer 2015; Sørensen 2015). The passage of time may heighten this effect as archaeologists try to squeeze new finds into very old typologies whose creators constructed by drawing boundaries through our "regions of ignorance" (paraphrasing Campbell 2007).

Yet thoughtfully constructed typologies remain indispensable for most kinds of analysis (Fowler 2017) because they are at the heart of comparisons. We can also use multiple typologies with distinct purposes in our research programs and, as noted elsewhere in this chapter, it is possible to use fuzzy classification to reflect uncertainties. Furthermore, there is no reason why case studies of individuals or particular circumstances cannot complement analyses that depend on typologies. They simply have different purposes.

3.9 Summary

- Archaeologists collect data into compilations that they structure through systematics (classification and grouping).
- Classification and grouping have different strengths and weaknesses in particular archaeological contexts, and typically we use both.
- Useful typologies make use of both intrinsic attributes of artifacts or sites and extrinsic (contextual) ones that plausibly relate categories to things like chronology, function, social groupings or networks, and technological practices.
- As with other kinds of observations, classifications have the potential for error, and protecting the validity of inferences based on them requires attention to quality control.
- It is also important to consider the implications of using rigid typologies with very old intellectual roots, as useful typologies have very specific purposes and should be responsive to new information.

References Cited

Adams, W. Y., & Adams, E. W. (1991). *Archaeological typology and practical reality, a dialectical approach to artifact classification and sorting*. Cambridge: Cambridge University Press.

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Beverley Hills: Sage.

Baxter, M. J. (1994). *Exploratory multivariate analysis in archaeology*. Edinburgh: Edinburgh University Press.

Baxter, M. J. (2003). *Statistics in archaeology*. London: Arnold.

Baxter, M. J. (2006). A review of supervised and unsupervised pattern recognition in archaeometry. *Archaeometry, 48*(4), 671–694.

Baxter, M. J. (2009). Archaeological data analysis and fuzzy clustering. *Archaeometry, 51*(6), 1035–1054.

Boozer, A. (2015). The tyranny of typologies: Evidential reasoning in Romano-Egyptian domestic archaeology. In R. Chapman & A. Wylie (Eds.), *Material evidence: Learning from archaeological practice* (pp. 92–109). London: Routledge.

Buck, C. E., Cavanagh, W. G., & Litton, C. D. (1996). *Bayesian approach to interpreting archaeological data*. New York: John Wiley & Sons.

Campbell, S. (2007). Rethinking Halaf chronologies. *Paléorient, 33*(1), 103–136.

Cann, J. R., Dixon, J. E., & Renfrew, C. (1969). Obsidian analysis and the obsidian trade. In D. Brothwell & E. Higgs (Eds.), *Science in archaeology: A survey of progress and research* (pp. 578–591). London: Thames & Hudson.

Childe, V. G. (1929). *The Danube in prehistory*. Oxford: Clarendon Press.

Childe, V. G. (1956). *Piecing together the past: The interpretation of archaeological data*. London: Routledge & Kegan Paul.

Clarke, D. L. (1968). *Analytical archaeology*. London: Methuen.

Coppi, R., Gil, M. A., & Kiers, H. A. L. (2006). The fuzzy approach to statistical analysis. *Computational Statistics and Data Analysis, 51*, 1–14.

Dellaportas, P. (1998). Bayesian classification of Neolithic tools. *Journal of the Royal Statistical Society C (Applied Statistics), 47*(2), 279–297.

Döring, C., Lesot, M.-J., & Kruse, R. (2006). Data analysis with fuzzy clustering methods. *Computational Statistics & Data Analysis, 51*(1), 192–214.

Drennan, R. D. (2010). *Statistics for archaeologists: A common sense approach*. New York: Springer.

Dunnell, R. (1971). *Systematics in prehistory*. New York: The Free Press.

Everitt, B. (1974). *Cluster analysis*. London: Heinemann.

Fowler, C. (2017). Relational typologies, assemblage theory and Early Bronze Age burials. *Cambridge Archaeological Journal, 27*(1), 95–109.

Gardin, J. C. (1980). *Archaeological constructs. An aspect of theoretical archaeology*. London: Cambridge University Press.

Garfinkel, Y. (1999). *Neolithic and chalcolithic pottery of the southern levant* (Qedem 39). Jerusalem: Hebrew University.

Gilboa, A., Karasik, A., Sharon, I., & Smilansky, U. (2004). Towards computerized typology and classification of ceramics. *Journal of Archaeological Science, 31*, 681–694.

Gonçalves, A. B., Souza, J. S., Da Silva, G. G., Cereda, M. P., Pott, A., Naka, M. H., & Pistori, H. (2016). Feature extraction and machine learning for the classification of Brazilian savannah pollen grains. *PLoS One, 11*(6), e0157044.

Goren, Y. (1992). Petrographic study of the pottery assemblage from Munhata. In Y. Garfinkel (Ed.), *The pottery assemblages of the Sha 'ar Hagolan and Rabah stages of Munhata (Israel)* (pp. 329–348). Paris: Association Paléorient.

Hall, M. (1982). Quantifying trends in site location with multidimensional scaling. *World Archaeology, 14*(1), 131–152.

Hand, D. J. (1997). *Construction and assessment of classification rules*. New York: Wiley.

Hodder, I. (1977). The distribution of material culture items in the Baringo district, western Kenya. *Man (N.S), 12*(2), 239–269.

Hodson, F. R. (1970). Cluster analysis and archaeology: Some new developments and applications. *World Archaeology, 1*, 299–320.

Hodson, F. R. (1982). Some aspects of archaeological classification. In R. Whallon & J. A. Brown (Eds.), *Essays on archaeological typology* (pp. 21–29). Evanston: Center for American Archaeology.

Jardine, C. J., Jardine, N., & Sibson, R. (1967). The structure and construction of taxonomic hierarchies. *Mathematical Biosciences, 1*, 173–179.

Kintigh, K. W. (1990). Intra-site spatial analysis in archaeology. In A. Voorrips (Ed.), *Mathematics and information science in archaeology: A flexible framework* (pp. 165–200). Borin: Holos-Verlag.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method* (2nd ed.). London: Butterworths.

Mackay, E. J. H. (1938). *Further excavations at Mohenjo-daro, being an official account of archaeological excavations carried out at Mohenjo-daro carried out by the government of India between the years 1927 and 1931*. New Delhi: Government of India.

Nance, J. D. (1987). Reliability, validity, and quantitative methods in archaeology. In M. S. Aldenderfe (Ed.), *Quantitative research in archaeology: Progress and prospects* (pp. 244–293). Newbury Park: Sage.

Needham, R. (1975). Polythetic classification: Convergence and consequences. *Man (N.S), 10*(3), 349–369.

Niccolucci, F., & Hermon, S. (2015). Time, chronology and classification. In J. Barcelo & I. Bogdanovic (Eds.), *Mathematics and archaeology* (pp. 257–271). Boca Raton: CRC Press.

Niccolucci, F., Hermon, S., & Doerr, M. (2015). The formal logical foundations of archaeological ontologies. In J. Barcelo & I. Bogdanovic (Eds.), *Mathematics and archaeology* (pp. 86–99). Boca Raton: CRC Press.

Orton, C. (1980). *Mathematics in archaeology*. Cambridge: Cambridge University Press.

Panagopoulos, M., Rousopoulos, P., Dafi, D., Papaodysseus, C., & Tracy, S. (2009). Automatic writer identification of ancient Greek inscriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(8), 144–1414.

Prentiss, W. C. (1998). The reliability and validity of a lithic assemblage typology: Implications for archaeological interpretation. *American Antiquity, 63*(4), 635–650.

Read, D. (1982). Towards a theory of archaeological classification. In R. Whallon & J. A. Brown (Eds.), *Essays in archaeological typology* (pp. 56–92). Evanston: Center for Amercian Archaeology.

Shennan, S. (1988). *Quantifying archaeology*. Edinburgh: Edinburgh University Press.

Sneath, P. H. A. (1962). The construction of taxonomic groups. In G. C. Ainsworth & P. H. A. Sneath (Eds.), *Microbial classification. 12th symposium of the Society for General Microbiology* (pp. 289–332). Cambridge: Cambridge University Press.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: Freeman.

Sørensen, M. L. (2015). 'Paradigm lost'—On the state of typology within archaeological theory. In K. Kristiansen, L. Smejda, & J. Turek (Eds.), *Paradigm found: Archaeological theory present, past and future* (pp. 84–94). Oxford: Oxbow.

Spaulding, A. (1953). Statistical techniques for the discovery of artifact types. *American Antiquity, 18*, 305–313.

Watson, P. J., LeBlanc, S., & Redman, C. (1971). *Explanation in archaeology, an explicitly scientific approach*. New York: Columbia University Press.

Weller, S. C., & Mann, N. C. (1997). Assessing rater performance without a "gold standard" using consensus theory. *Medical Decision Making, 17*, 71–79.

Weller, S. C., & Romney, A. K. (1988). *Systematic data collection*. Newbury Park: Sage.

Whallon, R. (1972). A new approach to pottery typology. *American Antiquity, 37*, 13–33.

Whittaker, J. C., Caulkins, D., & Kamp, K. A. (1998). Evaluating consistency in typology and classification. *Journal of Archaeological Method and Theory, 5*(2), 129–164.

Williams, L., Thomas, D. H., & Bettinger, R. (1973). Notions to numbers: Great Basin settlements as polythetic sets. In C. L. Redman (Ed.), *Research and theory in current archaeology* (pp. 215–237). New York: Wiley.

4 Compilations: Designing and Using Archaeological Databases

Bridging the gap between data and the information that data constitutes is perhaps the greatest challenge to the archaeological application of IT. . . . What we must avoid is unjustifiable 'black box' systems or we may find the discipline tarred with the familiar '. . . manipulation of ambiguous data by means of dubious methods to solve a problem that has not been defined'.

Cheetham and Haigh (1992: 13)

Archaeological research results in sometimes large "compilations," the products of collecting, compiling, cataloguing, or organizing archaeological data into what Gardin (1980) calls "arrangements." Compilations can be simple lists, indexed lists, computer databases, or online catalogues. Alternatively, they can be tables, graphs, "galleries" of pictures, or statistical summaries. One use for compilations is simply to record and summarize information, but we can also use them to organize that information in meaningful ways that help us understand the data. Compilations are increasingly at the forefront because of the demand for "big data" for certain research programs of great interest as well as the legal and ethical compulsions to preserve archaeological information in an accessible and useful form for future generations (Kintigh 2006). Compilations with a spatial focus are also extremely important in archaeology, which makes considerable use of Geographic Information Systems (GIS).

As noted in the previous chapter, the languages we use to describe data in any kind of compilation, or analyses and inferences based on the data, require an ontology—a system of concepts, categories, terms, or "things"—that allows us to describe and compare phenomena. In this chapter, you will see how we incorporate ontologies into a kind of language that is amenable to manipulation with computers.

The simplest kinds of compilations are just lists. Because they have no particular order, lists may not be the best way to manage or seek patterns in data unless we use software to help us search them. To speed up that process, it is often useful to have an index or keywords. Tables can also be useful to display data. Tables impose order and dimensions on the data and sometimes emphasize particular kinds of data over others. Graphs, meanwhile, simplify and make visual large quantities of data that we might otherwise have to present as almost unmanageable lists or tables of information. As they say, a picture is worth a thousand words. Further simplification is possible through statistical summaries, such as averages, medians and proportions. These replace lists of many numbers with only a few numbers that characterize trends, central tendencies, or typical distributions in the data.

Although compilations can be as simple as a collection of file cards, each with a picture of an artifact and some identifying labels and measurements, most modern archaeological compilations depend on digital information and computer databases. This section will introduce some principles that guide the effective use of these tools. For more general treatment of computers in archaeology, see Lock (2003) and Evans and Daly (2006).

4.1 Information Language

Useful compilations typically employ an information language that ensures consistent and efficient recording of observations. As discussed in Chap. 1, measurement in the broad sense consists of comparing an item with a standard scale and representing this symbolically. But natural language is usually too ambiguous, inconsistent or wordy to make these kinds of analyses consistently.

An information language is a system of representation. Although archaeologists have used various information languages for well over a century, the use of computers has led to more explicit attention to them. In the early days of computerization, archaeologists had to force themselves to

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_4

describe sites or artifacts with codes of no more than 80 characters. Today, we do not have so many constraints, but efficient and useful analyses still require the consistency of an information language, even if it sometimes resembles English or some other natural language.

4.1.1 Graphical Information Languages

Archaeologists have almost always used graphical conventions to simplify and make consistent the description of artifacts and other data. Maps and technical drawings are standardized simplifications of reality because they omit details that their makers or users do not find relevant while depicting or emphasizing information that would not be apparent in an unedited photograph or scan (see Chap. 21).

Take lithic illustrations, for example. Archaeologists less often publish photographs of lithics than drawings that encode important attributes of each tool, core or flake (Fig. 21.9). They use strict conventions or rules about the orientation of each piece, and how to depict ventral and dorsal views, side-by-side. The usual convention is to use solid lines to depict the borders of flake scars and curved, tapering lines to simulate the rippling that occurs in knapped flint to signal the direction of flake removals. They use stippling to indicate cortex and other conventions to represent non-flint materials, polish, and the position and direction of burin spall removals. The best lithic illustrations may seem like beautiful works of art, but they are actually simplified and standardized representations that emphasize the attributes that lithic researchers consider important and ignore or suppress others.

Pottery drawings are another excellent example of graphical information language. Ceramic illustrations usually bear little literal resemblance to the fragments of vessels on which they are based. Illustrators typically use information from the curvature of a sherd to estimate the diameter of the whole vessel, and then reconstruct as much as they can of that vessel from the fragmentary evidence available. In addition, they simultaneously depict the interior and exterior of the vessel and a radial cross-section through the vessel walls, sometimes called a "profile" (Fig. 21.10). The illustrator uses conventions for indicating any surface treatments and decorations that appear on the sherd, and often offers a cross-section through a handle, should there be one. Some information languages also use small icons and labels to encode additional information directly on the illustration (e.g., Smith 1973; Fig. 4.1).

While at first glance artifact illustrations may simply seem like drawings, in fact they are coded representations that display more information than would a photograph or even an unedited 3D digital model.

4.1.2 Digital Information Languages

While graphical information languages are excellent at depicting and storing information, they do not currently offer an efficient and accurate way to sort or retrieve or "query" that information, especially in large databases, although there are recent steps in that direction. Ideally, we want retrieval of information to be as easy as storing it.

Digital data processing uses an information language to reduce complex objects to a conventional representation in a database or to amplify a simple query into alternative forms that may represent it in the database (Gardin 1980). The former makes it easier to store large quantities of information in a way that allows us to retrieve it later; the latter allows us to ask questions of the database in order to make that retrieval.

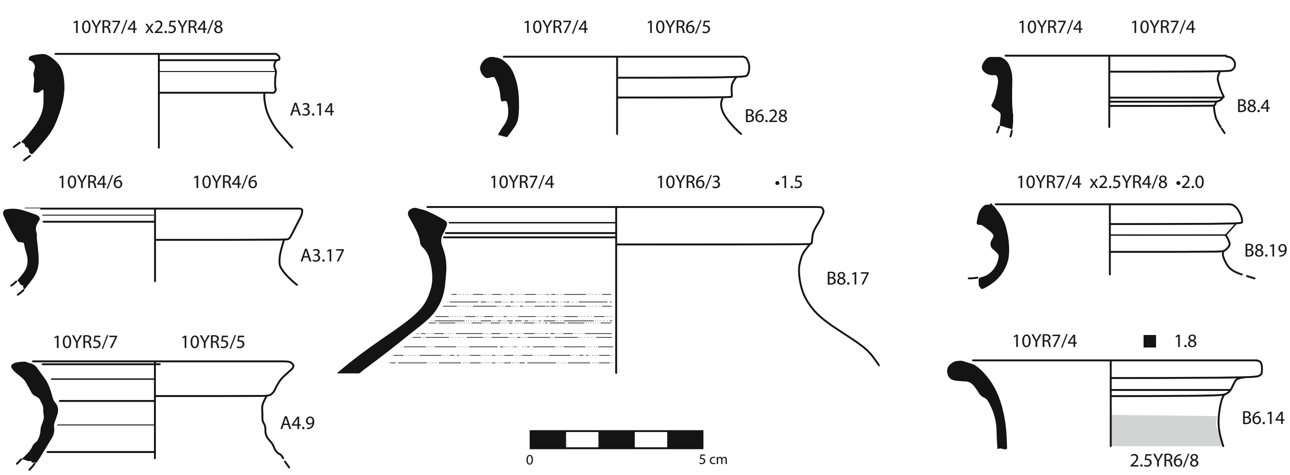


Fig. 4.1 An example of illustrations of pottery that employ a combination of graphical and textual information languages to display vessel shape and size, radial section, interior and exterior colors, surface treatments and artifact numbers. (cf. Smith 1973)

Information languages typically involve a specialized vocabulary, or **lexical units** (e.g., geometrical shapes, flake edges, decorative panels, brush strokes), **orientation rules** that define a standard position for the objects being described, **segmentation rules** for dividing an object into its constituent parts, and **differentiation rules** that specify the distinctions we record for each segment (Gardin 1980).

In terms of systematics, lexical units are the symbolic representations of attributes and categories, differentiation rules are related to the definition of categories and their separability, and rules for orientation and segmentation are other kinds of categorization that encourage consistency in the way we measure attributes. Because this involves rules, we are generally using classification, not grouping; the rules are definitions intended to make classes that are mutually exclusive and unambiguous, if possible. Although it is possible to make database queries that are more forgiving—for example, ones that would show you artifacts that are similar to the specific type you asked to see or that use fuzzy classification (Niccolucci et al. 2001)—most databases do not facilitate that, and even something as simple as a spelling error in a type name could lead to omission of important data from your query result.

We can describe pottery, for example, with an infinity of attributes (see Chap. 12 for some of these). For now, let us consider some candidates for basic lexical units, orientation rules, segmentation rules and differentiation rules for a pottery database. For segmentation rules, we might have to define very carefully what we mean by "rim," "neck," "shoulder," "body," "base," and "handle" (Fig. 4.2). Differentiation rules define how to assign each sherd to a segment, while other differentiation rules guide its classification as a particular kind of "rim" or "handle," etc. These rules also need to anticipate that some sherds will include all or portions of more than one segment, such as a rim with an attached handle, and your database will need a consistent and sensible way to accommodate these. The orientation rules for pottery usually require sherds to be "stanced" in the position it most likely had when part of a whole pot, standing vertically (usually with rim horizontal, and on top, base horizontal, at bottom). Body sherds and decoration often have less certain orientation, but the rules need to accommodate that uncertainty as well. The lexicon would usually include symbolic representations of categories for rim shape, decorative elements, mineral inclusions, and so on.

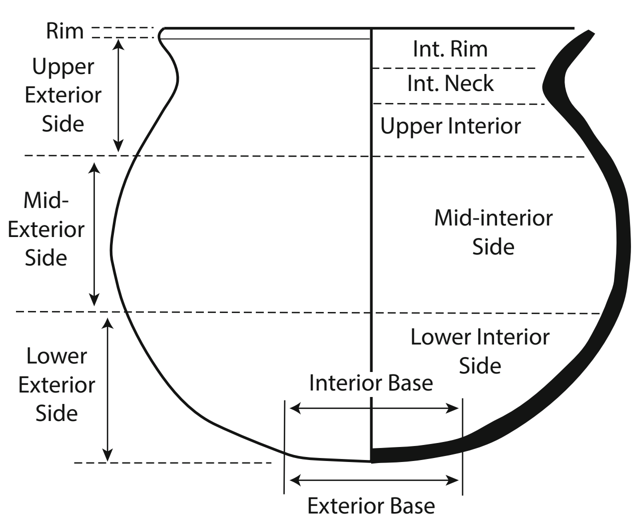


Fig. 4.2 An example of segmentation of a vessel into distinct parts (after Skibo 1992: 114). The definitions of these segments constitute the segmentation rules for an information language to describe pots like this

Ideally, the differentiation rules, like the definitions in a classification, allow unambiguous assignment of each sherd to a category. Often this involves a step-by-step, hierarchical procedure paralleling a taxonomic classification, but paradigmatic rules are also possible, and computers facilitate use of paradigms with many dimensions. Some kinds of data, such as rim shape, may be less amenable to unambiguous definition, and it is here that some archaeologists adopt a central-tendency form of grouping, with pictures of "ideal" rim shapes so that lab personnel can match each sherd to the shape to which it is most similar. On the other hand, modern databases may use mathematical characterizations of artifact shapes that allow stricter definition of shape classes (e.g., Gilboa et al. 2004, 2013).

Information languages for lithic materials typically employ such segments as "dorsal" and "ventral" sides and "proximal" and "distal" ends of flakes and have explicit orientation rules that define the axes of length and width and the position of retouch (see pp. 164, 168). The lexicon includes labels for categories of retouch, angles, raw material, platform shape and so on.

Decoration on pottery, basketry and other materials poses special problems. The orientation and segmentation rules often have to be complex to accommodate the distinctions that analysts find important and the way design elements are combined (e.g., Fig. 4.3). The alternative is a very lengthy lexicon that has separate categories for whole decorative schemes, rather than rules for the combinations of a smaller number of elements (e.g., Fig. 4.4).

Ancient Mesopotamian cylinder seals provide ambitious examples of archaeologists' attempts to describe artifacts with an information language. Digard et al. (1975) created a large lexicon for the symbols, figures, animals, buildings, furniture, and other things depicted on seals, and complex segmentation rules and a grammar or "syntax" that allows users to distinguish the different ways the lexical elements are combined and interact, including "actions," "number," and "configuration." For example, on a seal that shows a king, a

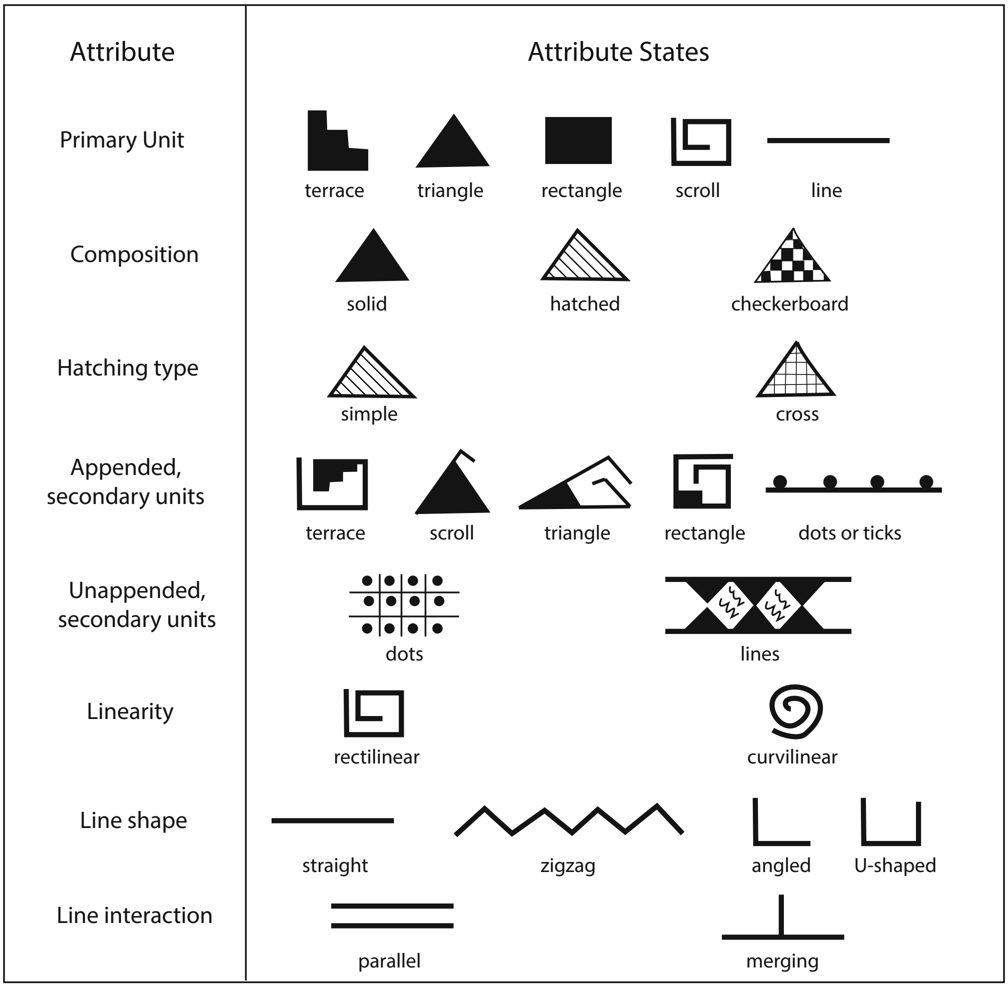


Fig. 4.3 An example of a portion of the differentiation rules for pottery decoration in the American Southwest. This one includes primary and secondary decorative elements that can be combined in various ways. (After Plog 1980: 51)

god, a throne, and a star, we can describe whether the king or god sits on the throne, whether the king is left or right of the god, is kneeling or standing, who is wearing which distinctive clothing, and the position of the star relative to other elements (Fig. 4.5).

Information language for describing architecture can focus on two-dimensional plans of structures or more thorough description. A database for the plans of rectilinear buildings, for example, needs orientation rules to identify things like long axis, short axis, front, sides and back of each building or room, as well as the azimuth (compass orientation) of the long axis. It would need lexical units to classify various rooms, features, post arrangements, stairs, or doorways (Fig. 4.6). It could also have a sort of grammar, like the one for cylinder seals, to describe the relationships between lexical elements, as Hillier and Hanson (1984) did for the "grammar" of built spaces. Other architectural lexical elements would likely include room dimensions, building area, and construction method.

4.2 Database Design

A database is a compilation of information that supplies users with data from which they make decisions and inferences, formulate typologies, explore patterns, or test hypotheses. It can be as simple as a telephone book or index card file, but today we are more likely to use the term to label an integrated body of digital information that we access through a computer or other digital device (even our phone). An integrated database has interrelated data stored with controlled redundancy, which means that only certain kinds of data are repeated, allowing us to retrieve other data that are unique. It can be the basis for complex data analysis, such as multivariate analysis of artifact similarities and differences, spatial analyses, or network analyses (Baxter 1994; Östborn and Gerding 2014).

Most databases share basic kinds of input and output. Users make inquiries, add or modify data (these are

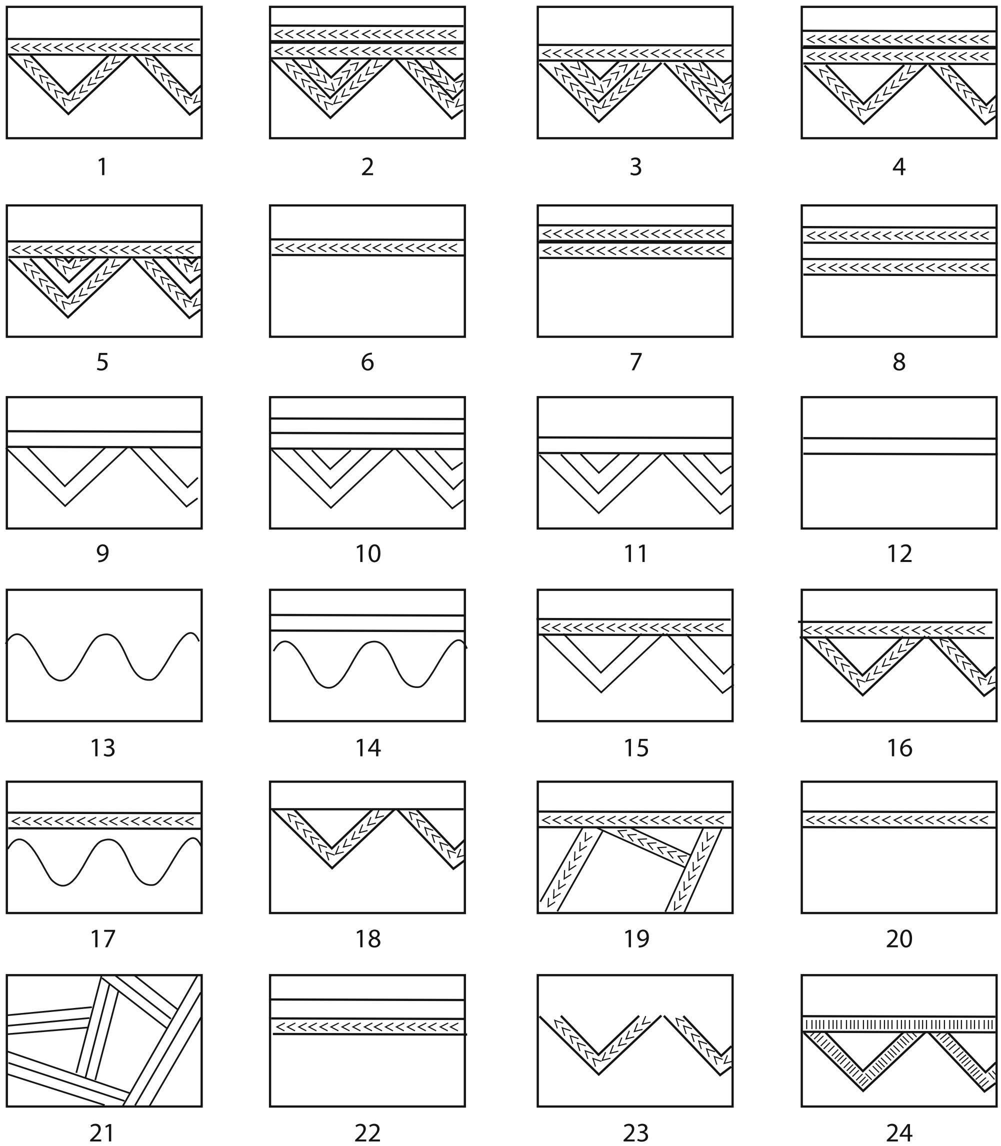


Fig. 4.4 An alternative lexicon for incised decoration on Yarmoukian Neolithic pottery that has a single level of design elements (after Garfinkel 1999: 65). Can you think of ways to redesign this as a sort of grammar with primary and secondary design elements and their interactions?

"transactions"), or they modify or add to the database itself. All these are kinds of input. For example, an archaeologist's input might consist of typing the description of a pit feature into one of the "fields" of a "record." Output consists of responses to inquiries, transaction logs (records of changes to the database), updated data, and an updated database. For example, the response to the query, "which pit features contained charred plant remains?" would trigger as output a list of the pit features that did.

Too often, archaeologists wanting to establish a database begin by sitting in front of a computer and defining some fields, without much forethought. This often results in a database that requires many revisions and corrections before it is even minimally acceptable for use, and this leads to frustration and sometimes much greater cost.

Just as with other aspects of research design (see Chap. 6), designing a database deserves better. It should begin with a list of objectives and expectations. Who will use the database, and for what purpose? Will its use be limited to a short time, or a single project, or do we expect it to serve for multiple projects for a long time? If the latter, who will maintain it and keep it up to date? What kinds of research questions will the database help to answer? Are the data types available in a commercial off-the-shelf database product sufficient for our purposes, or will we need to define specialized data types? Is there an existing database that is still useful and usable for our

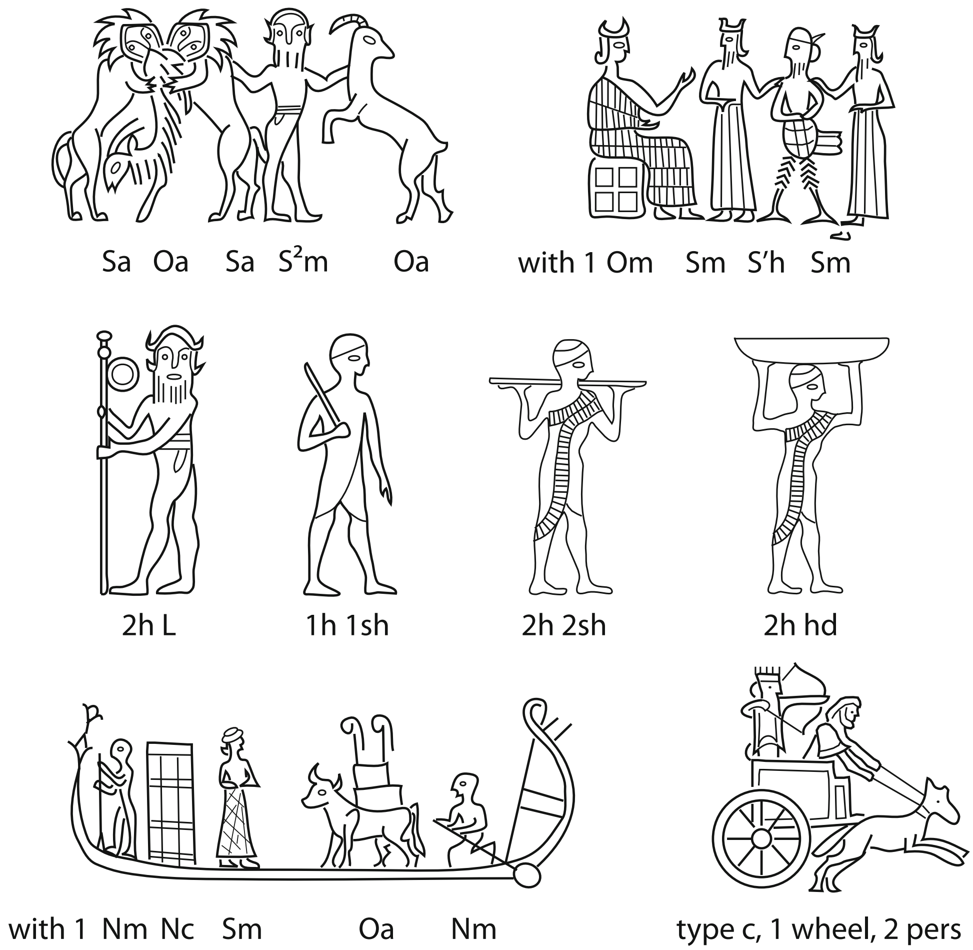


Fig. 4.5 A few examples of lexical items and interactions for describing cylinder seals. Top: interactions among people and animals. Middle: holding an object. Bottom: people, objects and animals and their interactions on vehicles. Lexical items include animal subject (Sa), animal object (Oa), subject man (Sm), object man (Om), hybrid creature being led (S^2h), neutral man (Nm), neutral container (Nc), and actions include holding two-handed to left (2 h L), one-handed on one shoulder (1 h 1sh), two-handed on both shoulders (2 h 2sh), and two-handed on head. (2 h hd; modified from Digard et al. 1975: 43, 45, 249, 325)

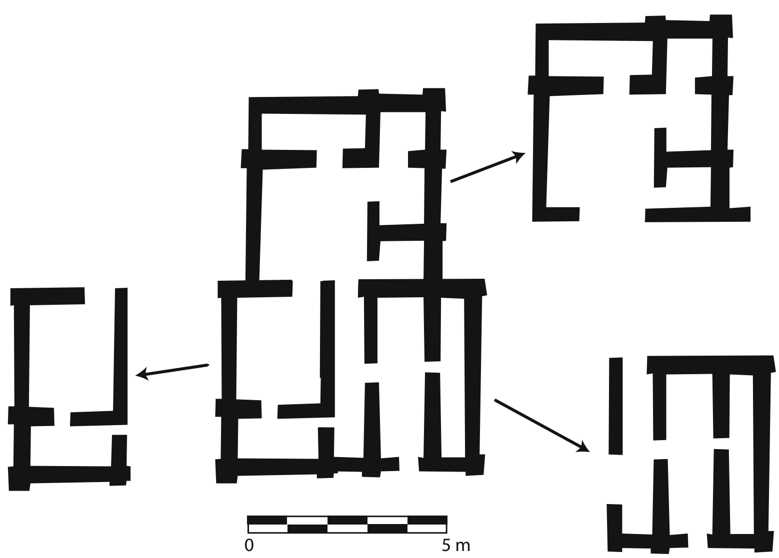


Fig. 4.6 A subset of segmentation rules for Samarran house plans in Mesopotamia, using the orientation rule that the widest part of the house is shown at the bottom. (After Banning 1997: 24)

purposes, or do we need to begin a new one from the ground up? If one already exists, what is the system like and what are its limitations?

Whether you are the principal or only user of the database you are designing or are collaborating with others, you should begin to create a logical design for the database before you even begin to think about its physical characteristics. Start by laying out how you expect the database to work and what kind of structure it should have to facilitate your anticipated uses rather than worrying too soon about what kind of processor, storage system, or software you should buy.

4.3 Data Models

To specify what you expect a proposed database to accomplish, it is useful to model it in the abstract. Start by listing the general kinds of entities (e.g., lithics, pottery, plant remains, users, sites, layers, photographs) that the database might need to document. An entity is any specific case of an entity type, so that a particular stone tool would be an entity potentially within the entity type, “lithics.” Then think about what kinds of attributes each of those entity types should have—size, shape, location, material, and so on—and any relationships that may exist among entities. For example, sites may contain layers or excavation units, pottery may have been found within layers, and a photograph may “capture” a view of a feature in a site. Relationships can also have attributes, such as the date when a layer “was excavated.” You might find it useful to think of entities as nouns, specific cases of an entity as proper nouns or I.D. numbers, and relationships as verbs (Chen 1976).

At the most general level, a conceptual data model outlines the main features and overall scope of your planned database. At a more practical level, you will want to have one or more logical data models that capture, in more detail, the entities already defined and any new entities you define to describe operations or transactions within the database, such as user logs or catalogs of attribute states (see Data Dictionaries, below).

Entity-relationship models (ERM) are graphical representations of these data models. There are several competing versions of ERM, but an example of one of these is in Fig. 4.7. These models identify the entities, relationships between entities, and attributes of both. For example, a “photo” entity’s attributes could include both a unique identifier and the date a photo was taken. The relationship, “supervises,” as when a person supervises excavation of a unit or a field crew, can have attributes like “date supervision began” or “directly or indirectly.”

ERMs usually specify two very specific attributes of relationships: whether the relationship is mandatory or optional, and whether it applies to one or multiple members of the entities it connects. In “crow’s foot notation,” a mandatory relationship is indicated by a vertical bar and an optional one by a circle, while relationships to a unique member of an entity is indicated by another vertical bar and those to multiple members by a “crow’s foot” or fork (Fig. 4.7).

4.4 Database Structure

Databases can be simple **flat-file databases**, analogous to a collection of file cards, more complex **relational databases** in which different files automatically communicate with one another, or **graph databases** (GBD) that use networks for very fast queries of highly interconnected data. **Query languages** are computer languages that allow users to retrieve and manipulate information in a database, including creating, modifying and deleting files or tables of data. The most popular query languages are versions of SQL (Structured Query Language).

Typical spreadsheets fall close to the flat-file end of this spectrum even though they may support some kinds of queries and searches. Graph databases are relatively new and no broadly accepted query language yet exists for them. However, they have great potential for archaeology given archaeologists’ growing use of network analysis as a tool for understanding a wide variety of spatial, social and historical phenomena (Mills 2017). As this chapter will only superficially refer to them, readers interested in graph databases should consult more specialized literature (Angles and Gutierrez 2008; Robinson et al. 2015; Yoon et al. 2017). Relational databases, which continue to play a very important role in archaeology, will be the focus of this chapter.

All databases consist of one or more files (sometimes called “tables” or, in the case of GDB, “nodes”), each of which corresponds to an entity type in a data model and has a number of fields and records. A **field** is a sort of abstract container for information on a particular attribute or characteristic of some item. For example, for pottery you might have a field for rim diameter in centimeters. A **record**

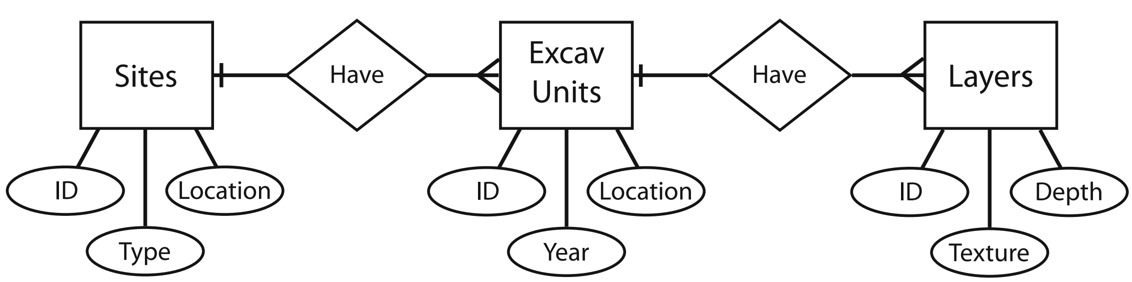


Fig. 4.7 Example of a portion of an Entity-Relationship Model (ERM) for a project with multiple sites and excavations. In this style of ERM, rectangles are entities or entity types, ovals are some attributes of those entities, and diamonds are relationships. On the connecting lines, the vertical bar at one end and “crow’s foot” at the other means that multiple members of the entity attached by the “crow’s foot” may participate in the relationship with only one member the other entity

(sometimes called a row or "tuple") is analogous to a single card in a card catalogue and corresponds to a row in a spreadsheet or a specific entity belonging to an entity type. It may describe a single site, artifact, feature, or some other phenomenon by displaying the contents of multiple fields. For example, a file to describe "sites" might contain 100 records for 100 different sites from an archaeological survey, while each record in that file describes a particular site with fields for "site number," "site size," "map coordinates," "elevation," "site type," and so on (Fig. 4.8).

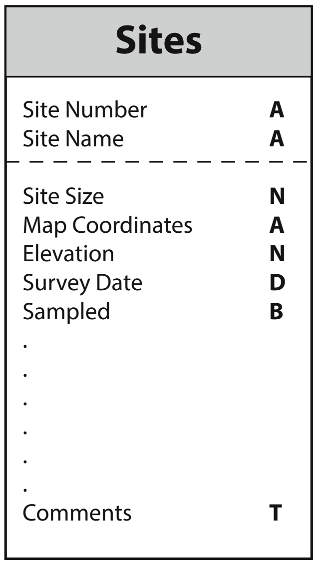
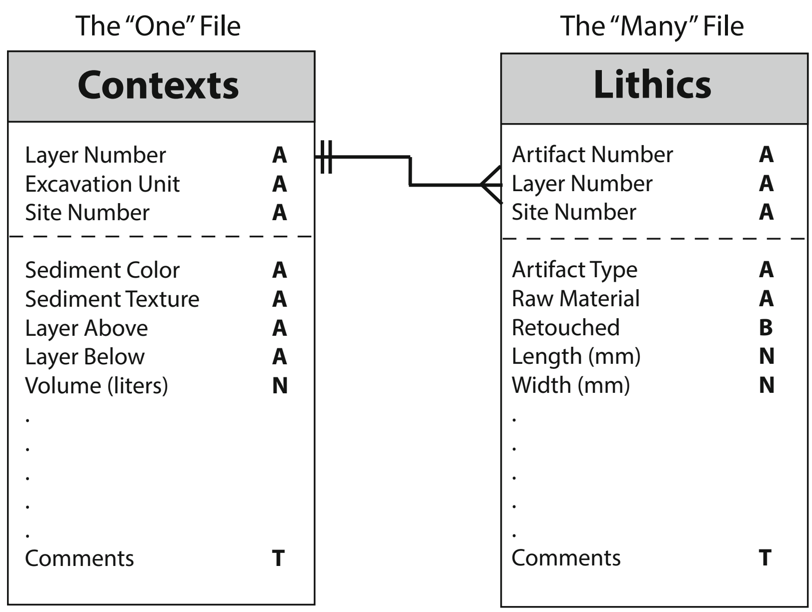


Fig. 4.8 A structure chart for a file to record "sites" information, with a key attribute (site number) and unique "site name" field above the dashed line, and a number of fields to describe other attributes of each site in a regional survey below. Bolded letters indicate whether the fields are alphanumeric (A), numeric (N), date (D), boolean (B), or text (T) fields. Note that "Site number" is *not* a numeric field because the numbers are arbitrary labels

A file for "Lithics," meanwhile, might have fields for "artifact number," "invasiveness of retouch," "number of retouched edges," "platform shape," or "segment" (Fig. 4.9). Many databases also define "forms" that dictate the way data will look during input and output, on your screen or in hard-copy. For most types of record, you will also have an input form and an output form that can mimic the paper forms that we once would have used to record data, or the reports that we would create with the data. In the case of spreadsheets, one kind of view, with rows for records and columns for fields, is used for both input and output. This is fine for many purposes, especially when there are only a few fields (columns), but is cumbersome for complex databases because users have to scroll around too much. Data entry is usually simpler with a form that devotes an entire page or screen to the data for a single record, mimicking the way we would fill out a paper form.

For large, modern archaeological projects, a simple flat-file database is unlikely to be very helpful, and relational databases' much richer, more flexible options (Johnson 2018) have made them the mainstay of archaeological data management. One of the things that makes relational databases preferable is "controlled redundancy," which allows us to make use of the relationships among classes of data, including spatial and stratigraphic context, in an efficient, hierarchical manner (Date 1986; Weinberg 1992).

Fig. 4.9 A structure chart depicting a relation between a "lithics" file and a "context" file for an archaeological excavation. Only some of the fields are shown. Bolded letters indicate whether the fields are alphanumeric (A), numeric (N), date (D), boolean (B), or text (T) fields. The "crow's foot notation" on the connecting lines indicate that each lithic has a mandatory relationship to a single member of the "context" file (double bar) while each context has an optional relationship to potentially many members of the "lithics" file ("crow's foot"). Note that "site number" in the context file could be an attribute pointer to a key attribute in a "sites" file as in Fig. 4.8

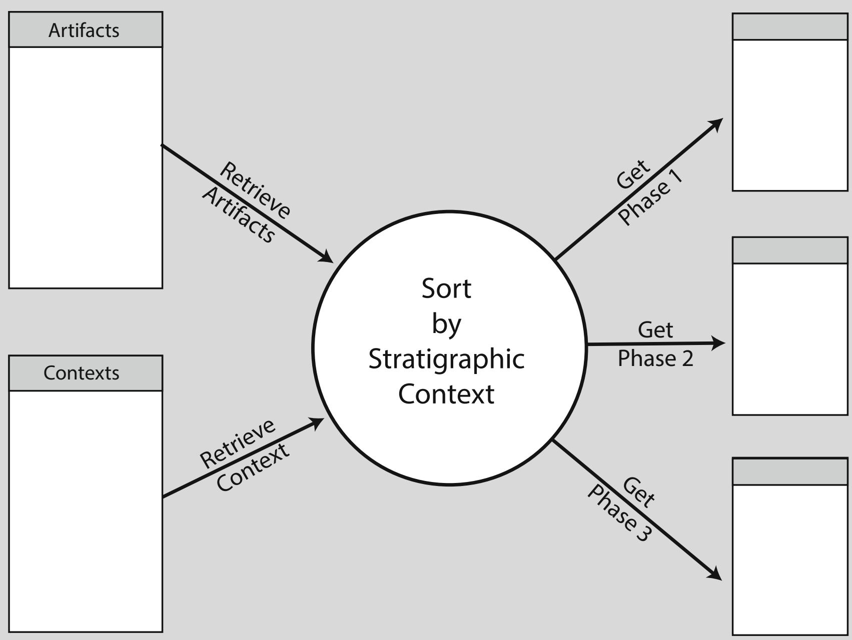


Example

Figure 4.10 illustrates how controlled redundancy works. If we are interested in the stratigraphic context of lithics in our database, we could put all the information into one big file, but that would be very inefficient. For one thing, many of the lithics would have nearly identical contextual information, such as sediment characteristics, dates excavated, and spatial boundaries. A flat-file structure would force us to enter that description of context on every record, in other words, with extreme redundancy. All of the lithics found in the same layer or pit at a site would be associated with the same sediment texture and color, so why waste time describing this potentially hundreds of times? In any case, it does not make much sense to record sediment attributes in a "lithics" file.

Instead, we should remember that lithics and contexts are different entity types and use different files for describing lithics and for describing the contexts in which they were found. We can then make a **relation** between them. The relation simply tells the computer where to go to look up the contextual information for whatever lithic record we are currently viewing or, in the other direction, to make a list of all the lithics that come from the same context when we are viewing a context record. It is possible to display information in the file we are viewing that actually comes from a completely different file with which it is linked, and to do this automatically. You do not have to leave the file and come back, let alone duplicate the information, as long as you set up the database accordingly. The relation also makes possible queries like "did we find any bone fragments in the same context as these lithics?" or "what kinds of bone fragments were associated with these lithics?" Similarly, we can have a relation between the "contexts" file and the "sites" file shown in Fig. 4.8.

Fig. 4.10 An example of a portion of a Data Flow Diagram (DFD) to sketch out what happens when someone carries out a process that sorts artifacts into groups on the basis of their stratigraphic context. For each artifact in the "artifacts" file, the process has to look up its stratigraphic placement by reference to the "context" file and then assigns it to one of the stratigraphic groupings at right



In a relational database, any relation consists of an **attribute pointer** (also called a "foreign key"), in this case a special field in the "Lithics" file, that points to a **key attribute** (or "primary key"), in this case in the "Contexts" file. Typically, the key attribute is just a name or ID number that uniquely identifies a record. In Fig. 4.9, both these fields have the same name, "Layer Number." In the "Contexts" file, there is only one record for each layer number (i.e., they are unique), and the other fields in each record describe the nature of the context (sediment character, stratigraphic and spatial position, etc.). In the "Lithics" file, by contrast, there could be many records with the same Layer Number, simply because many lithics were found in the same layer. For this particular situation, we would call the "Lithics" file the "Many File" and the "Contexts" file the "One File." Each unique record in the "One File" can have a relation with many records in the "Many File."

Figures 4.8 and 4.9 each show a **structure chart**, conventionally representing files as rectangles with the file name at the top and with the names and data types of fields below. The jagged line points from an attribute pointer to a key attribute, with a "crow's foot" at the "many" end of the relation and a bar at the "one" end. Some files can have multiple relations. For example, a "Contexts" file could have relations to "Sites," "Lithics," "Pottery," and "Photographs," among others.

4.4.1 Types of Fields (Data Types)

An **alphanumeric field** allows input of any characters typically available on a keyboard, including numerals, but is often limited to a specified number of characters. If you have defined an alphanumeric field with eight characters, the computer may store eight characters for each record even you do not input any data. Although today we worry a lot less about how much storage we use, you should still define alphanumeric fields to have only as many characters as you need. More sophisticated databases allow you to impose "filters" on alphanumeric fields. These help to prevent certain kinds of data-entry errors, such as inconsistent spelling, or allow standard entries to be selected from a drop-down menu, rather than typed. You can even specify, for example, that an alphanumeric entry must begin with two letters, or include some special character, or that a numeric one should have leading zeroes or never begin with "9." The content of an alphanumeric field consists of a *text string*.

Some databases have a version of an alphanumeric field called a **list field**. This makes it easier for you to define drop-down menus that ensure consistent spelling and limit data entries to preconceived categories of a classification.

Many database programs also have **text fields** or comment fields. These are alphanumeric but with no limitation on number of characters. They are ideal for situations where you might have little to say on some records, but a lot of information for others, potentially up to several pages, because they only take up as much storage space as you need; empty text fields have no cost. Text fields are excellent for logs and free-form descriptions, such as excavators' unscripted observations, providing the freedom to write richer, more nuanced prose than ordinary fields allow (cf. Hodder 1989).

A **numeric field** only allows users to input numerals. Most databases allow you to specify this data type more precisely, either with filters that allow you to limit entry to things that look like telephone numbers or GPS coordinates, or just by distinguishing integer from "real" (or "float") data types. Integer numerical fields are useful for discrete data, such as artifact counts, and, in most commercial databases, we need to select the "real" data type if we want a continuous scale. Both integer and "real" types allow negative numbers, and these databases do not distinguish interval from ratio scales, so the computer has no way to distinguish genuine zeroes from arbitrary ones unless you write a special program ("script" or "procedure") to make that distinction.

One great advantage of numeric fields is that many digital measuring instruments, such as electronic calipers and balances, can use an interface to input ratio-scale measurements directly into a database. This eliminates one potential source of error.

Decades ago, archaeologists had to "code" nominal-scale data with numerals instead of using alphanumeric labels in order to save storage space, because computers were slow and had little memory. Today, that is not necessary. Although some people still prefer to use codes, they should be careful because computer software can easily mistake numeric codes, including things like site numbers and layer numbers, for integer numbers; never put such codes in an integer field, as nominal data should always be alphanumeric, even when they look like numbers. However, if you sort numbers in an alphanumeric field, they will be sorted alphabetically, which means they will be out of order unless you use leading zeroes (e.g., "001," not "1").

Boolean fields are fields for dichotomous data. From the computer's perspective, they are like switches that can be on or off, and thus can register 1 or 0, true or false, yes or no. Archaeologists can also use these to represent other dichotomous observations, such as male/female in skeletons, above/below for stratigraphy, present/absent, left/right for skeletal elements. Convenient though it is, you should be careful that a Boolean field is really what you want before you lock it in; is it likely that you will have any "grey areas" or fuzzy classifications, such as "probably female" or "indeterminate"? If that is a possibility, you might be better to use a different data type that will not force you to make tedious "recoding" adjustments to your database after the fact.

Another field type that modern databases often offer is a **picture field**. This allows you to insert a photo or illustration into a record, but usually does not allow you to use characteristics of the picture as search operators, although Artificial Intelligence is rapidly changing that limitation. Picture fields, even in their non-searchable version, are useful for documenting artifacts and field contexts, alongside the other data in the record, as well as for maintaining photographic records. Digital photography makes it easy to take hundreds of field photos but, without such a record, it is equally easy to lose track of which photo documents what.

Archaeologists use many kinds of data that do not fit neatly into simple numeric or alphanumeric data types (Ryan 1992), or even picture fields. For example, a lot of archaeological data specifically involve the dimension of time as measured in radiocarbon years, calendar years, stratigraphic order, or cultural periods. If we treat information such as "Neolithic" just as alphanumeric data or enter "Layer 6" with a "6" in an integer or alphanumeric field, any attempt to order our data meaningfully is hopeless. The computer would sort the alphanumeric data alphabetically and the stratigraphy units either numerically or alphabetically with no reference to their true stratigraphic order. To solve this problem, we either have to write a procedure in the software to define the correct order of periods and stratigraphic units, or "trick" the software by coding them with a numerical prefix (e.g., "Neolithic" might become "06 Neolithic" to ensure that it comes after "05 Epipalaeolithic" and before "07 Chalcolithic"). In digital information languages, what we want here is an **abstract data type**, in part because we sometimes need distinctions, not just between "Neolithic" and "Chalcolithic," but between "Chalcolithic" and "Late Chalcolithic." Archaeologists frequently need such flexible distinctions.

Among the abstract data types that commercial database software routinely includes are **date fields**, which allow you to enter data in day/month/year or month/day/year or year/month/day formats and which sorts the data chronologically instead of numerically. Date fields also allow you to use the search operators "earlier than," "later than" and "contemporary with." This can be handy, for example, when you want to separate out all the excavation units that were excavated in 2009 and 2010, or to find all the records that were entered before October 10, 2016. The latter is an example of a **transaction time** (the time when someone entered or modified a record), while the former is called a **valid time**. Most databases allow us to have data entries automatically "date-stamped" with a transaction time. However, the date fields in commercial software are insufficient for archaeologists' needs because they do not easily accommodate "fuzzy" dates like "Archaic" and "Late Archaic" or dates with errors and confidence intervals, such as "1050 ± 150 BP" or "1257-844 cal BC at 95% credible interval." Archaeologists need special temporal operators that supplement ones like "earlier than" by allowing overlaps and statements of confidence and probability (Cheetham and Haigh 1992). Ideally, for example, we should be able to search databases for records that date within one standard deviation of 5000 BP, what Cheetham and Haigh (1992: 12) call a **statistical date type**. Even now, archaeologists need to write scripts (small programs) to accomplish this.

Other than chronological ones, the next most important abstract data types for archaeology are spatial. These include things like a site's GPS coordinates, spatial coordinates within sites, the shapes of artifacts, and the locations of cut-marks on bones. For example, we might want to query a database to identify all artifacts found within 2 m of a particular feature or all photographs in our photographic archive that include a particular point on a site (Ryan 1992: 5). Today, archaeologists accommodate most of their spatial needs at the level of sites and landscapes by using Geographic Information Systems (GIS). These are specifically spatial databases (see Conolly and Lake 2006; Lock and Pouncett 2017; Scianna and Villa 2011) and do not need to be limited to large scales.

Besides date fields, commercial databases offer abstract data types for hours/minutes/seconds, money, and telephone numbers, mostly with little archaeological application. Because their main markets are in the commercial sector, software companies have not been quick to accommodate the abstract data types that archaeologists would require (Cheetham and Haigh 1992: 7).

Rapid growth in the use of touch-screen devices, especially tablets and phones, has provided hardware options that further expand the options for entering data. While databases running on these devices continue to use the kinds of fields described above, it is now easier to make entries by way of "buttons." Not only can we configure these as checkboxes on a form, we can overlay buttons on graphic images, such as maps or drawings. This has some advantages over drop-down menus. For example, rather than having to remember what term to use to describe the location of a cut-mark on a pig femur, we can tap the cut-mark's location on the appropriate surface on an image of a pig's femur on the screen (Dibble 2015). This is likely faster and arguably more accurate than typing, drop-down menus, or even checkboxes.

4.4.2 Operations or Procedures

The kinds of operations that you can apply to data depend on data type. Numeric fields are susceptible to all the usual arithmetical and appropriate (and sometimes inappropriate) statistical functions, including the operations "greater than," "sum," "product," "square root," and "average." For alphanumeric and text fields, you can use such operations as "contains," and "does not contain," much as in online searches. For example, you could ask the computer to show you all the records in which the field, "Site Name," contains the string (sequence of characters), "Koster" but whose field, "Flotation Results," contains "" (i.e., is empty). This is called a *search*. The computer would return all the records from the Koster Site for which the flotation results had not yet been entered. Another common operation for alphanumeric fields is to *sort* the records by one or more attributes.

A stored procedure is one that is saved in the database itself. This is useful for operations that we use so routinely that we want to automate them. For example, a stored procedure in a database for an archaeological survey using tablets for data entry might take data from fields that store GPS coordinates for the start and ending locations of each survey transect to calculate distance walked, and then use that and the number of artifacts recorded in another field to generate artifact density automatically (Banning and Hitchings 2015). The procedure then saves the result (an indirect measurement) in a field that some databases call a "calculated field."

4.4.3 Data Flow Diagrams

Data Flow Diagrams (DFDs) are one of the basic tools of Structured Analysis and Structured Design in computer science (Weinberg 1992). A DFD provides a logical model of an information system, no matter what physical form it takes, by showing its logical processes and flows of information. It is useful for modelling how you and others will make use of data in the planning stage of your database, to ensure that you set up that database in a way that facilitates, rather than frustrates, that work. It can be very tempting to sit at a computer and begin setting up a database without planning it ahead of time, but you should avoid this temptation. A DFD is a very useful tool to help you think about how users will want to interact with data so that you can plan it with these constraints in mind.

A DFD shows processes (the activities that users carry out with or on the data), flows (movement of data between processes or from files to processes to files or output), and data storage entities (files and reports). Typically, an arrow represents a flow of data, a labelled circle represents a process, and a labelled rectangle (as in the structure chart) represents a file or storage device.

A DFD allows you to sketch out the kinds of things you would like to do with your data. It is likely easier to start with simple parts or individual processes, rather than trying to model a whole project at the outset. For example, most likely you will need to retrieve information about artifacts that are sorted by some time dimension. One way to do that (Fig. 4.10) might involve a stratigraphic sorting that pulls information from both "Artifacts" and "Context" files and processes the data to send it to different files or reports that each include artifacts from a particular stratigraphic phase or group of phases.

Among the reasons to draw DFDs is that they help to highlight what kinds of data you need to retrieve, how often, how quickly, and with what level of detail. You may want to design your database very differently depending on whether certain kinds of information need to be retrieved daily, or only rarely. The DFD might also help you decide whether or not you need to develop an abstract data type for some aspect of the information, rather than making do with more conventional data types or "tricking" the computer with carefully designed codes, as you might do to ensure that stratigraphic order is sorted correctly. It will also highlight such things as what statistical tests, if any, you might want to make. Given that some tests have strict requirements and assumptions, that could make a big difference about how you measure things, let alone how you record them in your database. DFDs make it easier to identify potential problems before you start building the database or, worse yet, enter a lot of data you have to fix later.

4.5 Data Dictionaries and Metadata

Databases can quickly become complex, and a **data dictionary** is essential to document its components and how they work so that others, and even you, will not have to waste valuable time figuring it out later. A data dictionary (or catalog) documents the database's information language: the lexical units (files, fields, attributes, values, data flows) and database structure (relations, indices, attribute pointers) and any other aspects of the database that are relevant to its use, maintenance, and preservation for the future. These are the database's **metadata**: data about other data.

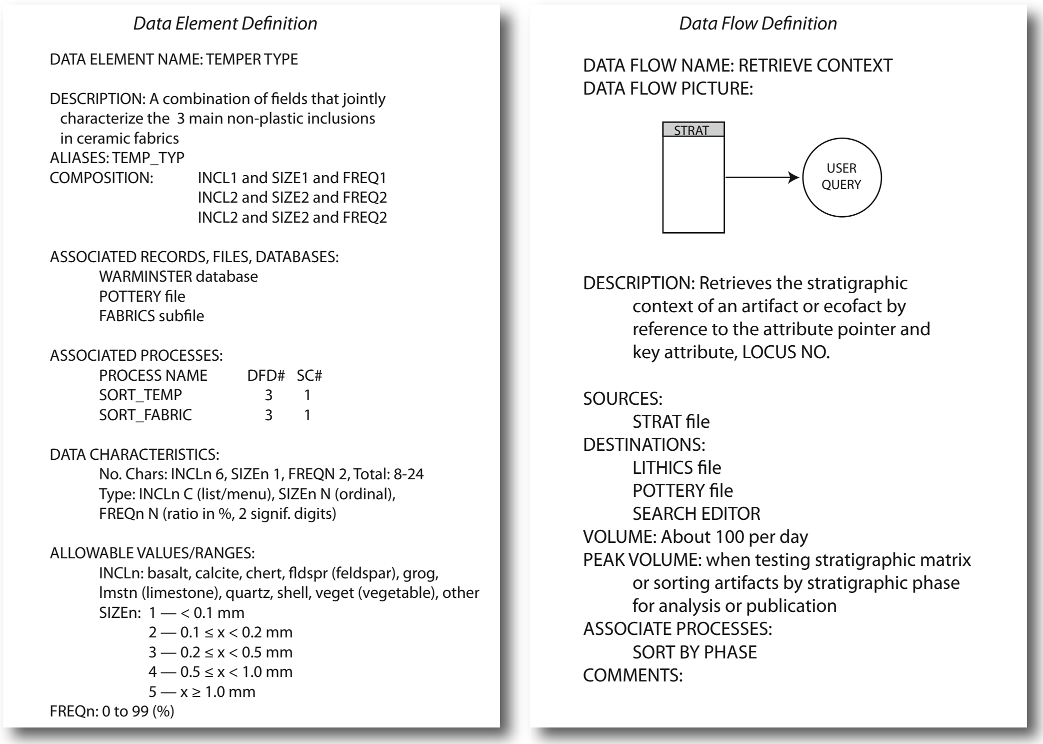
Although there are many ways you could set up a data dictionary, at a minimum it should document all the database's entities (files or tables and fields or columns) and relations among them, but it makes sense to document at least the most important data flows and logical processes. For files, it should include a structure chart and describe any associated relations, and for fields it should document the data type, scale of measurement, units, if any, filters or restrictions on character length, and the categories of any classification or list (e.g., drop-down menu). It should tell users where to find things (e.g., which files have a particular field), what their purpose is, and how and how often they are used (e.g., Fig. 4.11).

4.5.1 Metadata Challenges and Digital Curation

Over the cycle of a research project and beyond, software and hardware change substantially, file formats become obsolete, different researchers use different ways to structure data, and projects eventually come to an end. These can result in major challenges unless we ensure the preservation and curation of metadata to allow current and future researchers to access and use the data without sacrificing its integrity (Johnson 2018; Kulasekaran et al. 2014; Schmidt 2001). For two decades, there has been a trend to accumulate, disseminate and preserve the data from multiple archaeological projects, for example in the Digital Archaeological Record (tDAR), the Alexandria Archive Institute's Open Context, and the Archaeology Data Service (ADS) (Kansa and Kansa 2007; Niven 2013; Richards 1997; Sheehan 2015; Watts 2011). To enable "Big Data" analyses of such data, Holdaway et al. (2018) suggest keeping digital syntax simple, maintaining a focus on the relationships between basic archaeological phenomena, and placing a great deal of emphasis on offering detailed metadata. Yet many "Big Data" projects instead employ prescriptive metadata in an attempt to absorb competing data entities and structures within higher-level abstractions. More generally, contested ontologies and the proliferation of data-collection outside of "project" databases, such as team-members' and students' personal databases and even data captured on fieldworkers' phones, has led to something like a "Wild West" of digital data that prescriptive protocols for metadata are poorly equipped to incorporate (see also Dallas 2015, 2016; Kintigh et al. 2015). Yet another problem is that the curation of such data does not necessarily lead to its future use (Huggett 2018).

4.6 Web-Based Archaeological Databases

The need to preserve archaeological data as well as the desirability of sharing data among colleagues has led to the creation of large, online databases that include data from multiple research teams. Among the organizations that pursue this are Digital Antiquity, which hosts the digital archive, tDAR (https://core.tdar.org/; Sheehan 2015), the Online



Data Element Definition

DATA ELEMENT NAME: TEMPER TYPE

DESCRIPTION: A combination of fields that jointly
characterize the 3 main non-plastic inclusions
in ceramic fabrics
ALIASES: TEMP_TYP
COMPOSITION: INCL1 and SIZE1 and FREQ1
INCL2 and SIZE2 and FREQ2
INCL2 and SIZE2 and FREQ2

ASSOCIATED RECORDS, FILES, DATABASES:
WARMINSTER database
POTTERY file
FABRICS subfile

ASSOCIATED PROCESSES:

PROCESS NAME	DFD#	SC#
SORT_TEMP	3	1
SORT_FABRIC	3	1

DATA CHARACTERISTICS:
No. Chars: INCLn 6, SIZEn 1, FREQN 2, Total: 8-24
Type: INCLn C (list/menu), SIZEn N (ordinal),
FREQn N (ratio in %, 2 signif. digits)

ALLOWABLE VALUES/RANGES:
INCLn: basalt, calcite, chert, fldspr (feldspar), grog,
lmstn (limestone), quartz, shell, veget (vegetable), other
SIZEn: 1 — < 0.1 mm
2 — $0.1 \le x < 0.2$ mm
3 — $0.2 \le x < 0.5$ mm
4 — $0.5 \le x < 1.0$ mm
5 — $x \ge 1.0$ mm
FREQn: 0 to 99 (%)

Data Flow Definition

DATA FLOW NAME: RETRIEVE CONTEXT
DATA FLOW PICTURE:

DESCRIPTION: Retrieves the stratigraphic
context of an artifact or ecofact by
reference to the attribute pointer and
key attribute, LOCUS NO.

SOURCES:
STRAT file
DESTINATIONS:
LITHICS file
POTTERY file
SEARCH EDITOR
VOLUME: About 100 per day
PEAK VOLUME: when testing stratigraphic matrix
or sorting artifacts by stratigraphic phase
for analysis or publication
ASSOCIATE PROCESSES:
SORT BY PHASE
COMMENTS:

Fig. 4.11 Example of two pages from a data dictionary, one defining a data element, the other a data flow

Cultural and Historical Research Environment (OCHRE) at University of Chicago (https://voices.uchicago.edu/ochre/), and Open Context (https://www.opencontext.org/k). There are also more specialized online databases. For example, for radiocarbon data, there are BANADORA (https://www.arar.mom.fr/banadora/), the Canadian Archaeological Radiocarbon Database (CARD; Gajewski et al. 2011), CONTEXT (http://context-database.uni-koeln.de/), New Zealand Radiocarbon Database (https://www.waikato.ac.nz/nzcd/), Oxford University Radiocarbon Database (https://c14.arch.ox.ac.uk/login/login.php?Location=%2Fdatabase%2Fdb.php), and Radiocarbon Dates Online (RADON; https://radon.ufg.uni-kiel.de/).

4.6.1 Hypermedia and Online Data

Hypermedia are media, such as web sites, that use preconceived links in their text and images so that readers can jump from place to place and follow their own interests instead of reading linearly and sequentially, as in a novel, or doing searches and making queries. Some of the links can be to videos or 3D models of artifacts. However, unlike doing searches in a typical relational database, users of hypermedia follow links that an author has embedded in the media. In other words, the author can encourage readers to find more detail on another page or web site, where there could be still more links to other related information, while users of a relational database can make queries or search for information in ways that the database author did not anticipate, within the limitations of how the author or database manager structured the data. Although hypertext can occur in other media, today it is most common in web sites and social media. Hypermedia have much potential for the dissemination of complex archaeological information, while also presenting challenges with regard to making that information permanently available (see "conservation of digital information," p. 196).

4.7 Artificial Intelligence (AI), Data Mining, and Big Data

Today, AI increasingly allows pattern recognition that can mimic some of the classification and grouping processes of humans, while some archaeologists have been finding ways to "mine" data from multiple projects that used different data structures. "Big Data" refers to the fact that the resulting data sets are so large and complex that typical data-processing and statistical methods are not up to the task of interpreting them (Cooper and Green 2016).

For example, much archaeology today concerns Geospatial Big Data (GBD), which takes information from large numbers of GIS, remote sensing, excavation, radiocarbon and landscape-archaeology datasets in attempts to solve big questions, like the spread of domesticated livestock across SW Asia and Europe (Conolly et al. 2011), or cycles of demographic expansion and contraction in late prehistoric Ireland (McLaughlin et al. 2016; see also Chap. 20). Aside from other data-comparability issues, incorporation of older spatial data needs to account for pre-GPS problems with spatial precision and accuracy (Ullah 2015). Other "Big Data" initiatives attempt to integrate data from multiple projects that used somewhat different ways to structure their data by finding common vocabularies to make the data commensurable (e.g., Harrison 2019).

This is a difficult task. As noted above, computers normally expect us to define categories and terms very carefully because they will otherwise make serious mistakes, even in such simple things as treating "grey" as different from "gray." So how do we solve big problems with data in different databases, with different information languages and data structures that are not easily comparable? How do we make comparable datasets that employed different conceptual categories, sometimes quite divorced from the raw data on which those were based (Holdaway et al. 2018; Kintigh 2006)?

We can summarize some of the major challenges of such initiatives by the "seven Vs" (McCoy 2017). Volume is the problem that the size of datasets is so great that we talk, not in gigabytes or terrabytes, but in petabytes (millions of gigabytes). Velocity has to do with how quickly new data are being added, so it is hard for digital archives to keep up. Variety is the problem that different source projects used different conceptualizations, information languages and data structures, so it is difficult to tell whether two projects who use the same descriptors are talking about the same thing, or if different terms are really describing different things. Veracity is the problem of quality in the source data; some projects, especially from long ago, had poor precision in their spatial information, for example, and we need to correct this (Ullah 2015). Visualization has to do with the way we represent the data and identify patterns in it. Visibility, finally, concerns the increasingly high availability of data through web sites, cloud repositories, and gazetteers, but presents potential problems with privacy and misuse of data.

Despite these problems, AI has great potential for tackling problems that are too large for individual, locally based projects, as well as for making use of "legacy" data sets that might otherwise languish in project archives (Bevan 2012).

4.8 Quality Issues

A nagging challenge is that entry of data into a formal database may provide an illusion of certainty that is not really warranted. The mere fact that data entry involves having to check boxes, select from drop-down menus, tap specific areas on a touch screen, and fit observations into pre-established fields forces users to make decisions that may obscure real uncertainty or "fuzziness" in the original observations (Niccolucci et al. 2001: 3). Many of the potential objections to typologies (p. 40), are equally applicable to databases (Baines and Brophy 2006). While providing a "comments" field to accommodate extra information and concerns about the rigidity of the database structure may help, it is fairly obvious that most analysts who later make use of the database are likely to ignore these caveats. A possible response to this problem is to use fields that accommodate fuzzy measurements (e.g., indicating 80% confidence) instead of yes-or-no responses.

In addition, archaeological databases share with more old-fashioned compilations the potential for error in data entry, interpretation of terms, and use of data for purposes not originally envisioned. Often, digital and touch-screen data entry will perform better than, for example, handwritten notes and transcription (Dibble 2015), but they are not immune to error either. One reason to use drop-down menus, "filters," touch-screen buttons, and other features of database software that standardize data entry—despite the problem mentioned in the previous paragraph—is that they help to minimize nonsense errors like misspellings or confusion of "7" for "1" or "2."

However, they will not eliminate them entirely, and quality protocols that include random audits and checks for outliers in data that may be due to human error have an important role. While it is impossible to prevent inappropriate future uses of a database, having properly maintained documentation, such as a data dictionary, that is available to users will help to reduce the incidence of misapplications. Evaluating the effectiveness or success of a data system is by reference to targets or goals. These can include, besides the reduction of errors like those just mentioned, the costs of maintaining quality and how those compare with the costs of data errors and the costs of correcting those errors. Since the cost to correct or "clean" data is usually much higher than the cost of getting things right in the first place, these evaluations are essential.

4.9 Summary

- Relational databases are useful tools for managing compilations but require careful attention to the compilation's purpose and likely users
- Systematics is an essential component of database design, since databases use an information language to describe things as unambiguously as possible
- It is important to design a database system carefully "on paper" before beginning to create it in a software platform, usually with tools like entity-relationship models and data flow diagrams

- Databases employ information languages intended to ensure consistency of records
- Relational databases use controlled redundancy to record data efficiently while allowing access to information in related files
- Documenting all aspects of the database carefully is an essential component of the database's quality and future usefulness

References Cited

Angles, R., & Gutierrez, C. (2008). Survey of graph database models. *ACM Computing Surveys, 40*(1), 1–39.

Baines, A., & Brophy, K. (2006). What's another word for thesaurus? Data standards and classifying the past. In T. L. Evans & P. Daly (Eds.), *Digital archaeology: Bridging method and theory* (pp. 210–221). London: Routledge.

Banning, E. B. (1997). Spatial perspectives on early urban development in Mesopotamia. In W. E. Aufreht, N. A. Mirau, & S. W. Gauley (Eds.), *Aspects of urbanism in antiquity: From Mesopotamia to Crete* (pp. 17–34). Sheffield: Sheffield Academic Press.

Banning, E. B., & Hitchings, P. (2015). Digital archaeological survey: Using iPads in archaeological survey in Wadi Quseiba, northern Jordan. *The Archaeological Record, 15*(4), 31–36.

Baxter, M. J. (1994). *Exploratory multivariate analysis in archaeology*. Edinburgh: Edinburgh University Press.

Bevan, A. (2012). Spatial methods for analysing large-scale artefact inventories. *Antiquity, 86*(332), 492–506.

Cheetham, P. N., & Haigh, J. G. B. (1992). The archaeological database—New relations? In G. Lock & J. Moffett (Eds.), *Computer applications and quantitative methods in archaeology, 1992* (BAR international series S577) (pp. 7–14). Oxford: Tempus Reparatum.

Chen, P. (1976). The entity-relationship model: Toward a unified view of data. *ACM Transactions on Database Systems, 1*(1), 9–36.

Conolly, J., & Lake, M. (2006). *Geographical information systems in archaeology*. Cambridge: Cambridge University Press.

Conolly, J., Colledge, S., Dobney, K., Vigne, J. D., Peters, J., Stopp, B., Manning, K., & Shennan, S. (2011). Meta-analysis of zooarchaeological data from SW Asia and SE Europe provides insight into the origins and spread of animal husbandry. *Journal of Archaeological Science, 38*(3), 538–545.

Cooper, A., & Green, C. (2016). Embracing the complexities of 'big data' in archaeology: The case of the English landscape and identities project. *Journal of Archaeological Method and Theory, 23*(1), 271–304.

Dallas, C. (2015). Curating archaeological knowledge in the digital continuum: From practice to infrastructure. *Open Archaeology, 1* (1), 176–207. https://doi.org/10.1515/opar-2015-0011.

Dallas, C. (2016). Digital curation beyond the "wild frontier": A pragmatic approach. *Archival Science, 16*(4), 421–457.

Date, C. J. (1986). *Relational database: Selected writings*. London: Addison-Wesley.

Dibble, W. F. (2015). Data collection in zooarchaeology: Incorporating touch-screen, speech-recognition, barcodes, and GIS. *Ethnobiology Letters, 6*(2), 249–257.

Digard, F., Abellard, C., Bourelly, L., Deshayes, J., Gardin, J. C., leMaître, J., & Saloné, M. R. (1975). *Répertoire analytique des cylindres orientaux* (Vol. 2., *Code*). Paris: CNRS.

Evans, T. L., & Daly, P. (Eds.). (2006). *Digital archaeology: Bridging method and theory*. London: Routledge.

Gajewski, K., Muñoz, S., Peros, M., Vialu, A., Morlan, R., & Betts, M. (2011). The Canadian archaeological radiocarbon database (CARD): Archaeological ^{14}C dates in North America and their Paleoenvironmental context. *Radiocarbon, 53*(2), 371–394.

Gardin, J. C. (1980). *Archaeological constructs. An aspect of theoretical archaeology*. London: Cambridge University Press.

Garfinkel, Y. (1999). *Neolithic and chalcolithic pottery of the southern Levant*. Jerusalem: Institute of Archaeology, Hebrew University of Jerusalem.

Gilboa, A., Karasik, A., Sharon, I., & Smilansky, U. (2004). Towards computerized typology and classification of ceramics. *Journal of Archaeological Science, 31*, 681–694.

Gilboa, A., Tal, A., Shimshoni, I., & Kolomenkin, M. (2013). Computer-based, automatic recording and illustration of complex archaeological artifacts. *Journal of Archaeological Science, 40*, 1329–1339.

Harrison, T. P. (2019). Computational Research on the Ancient Near East (CRANE): Large-scale data integration and analysis in Near Eastern archaeology. *Levant, 51*(1), 1–4. https://doi.org/10.1080/00758914.2018.1492784.

Hillier, B., & Hanson, J. (1984). *The social logic of space*. Cambridge: Cambridge University Press.

Hodder, I. (1989). Writing archaeology: Site reports in context. *Antiquity, 63*, 268–274.

Holdaway, S. J., Emmit, J., Phillipps, R., & Masoud-Ansari, S. (2018). A minimalist approach to archaeological data management design. *Journal of Archaeological Method and Theory, 26*(2), 873–893. https://doi.org/10.1007/s10816-018-9399-6.

Huggett, J. (2018). Reuse remix recycle: Repurposing archaeological digital data. *Advances in Archaeological Practice, 6*(2), 93–104.

Johnson, I. (2018). Sustainability = separation: Keeping database structure, domain structure, and interface separate. In M. Matsumoto & E. Uleberg (Eds.), *CAA2016, Oceans of data. Proceedings of the 44th conference on computer applications and quantitative methods in archaeology* (pp. 63–68). Oxford: Archaeopress.

Kansa, S. W., & Kansa, E. C. (2007). Open content in Open Context. *Educational Technology, 47*(6), 26–31.

Kintigh, K. W. (2006). The promise and challenge of archaeological data integration. *American Antiquity, 71*(3), 567–578.

Kintigh, K. W., Altschul, J. H., Kinzig, A. P., Limp, W. F., Michener, W. K., Sabloff, J. A., et al. (2015). Cultural dynamics, deep time, and data planning cyberinfratructure investments for archaeology. *Advances in Archaeological Practice, 3*, 1–15.

Kulasekaran, S., Trelogan, J., Esteva, M., & Johnson, M. (2014). Metadata integration for an archaeology collection architecture. In W. E. Moen & A. Rushing (Eds.), *Proceedings of the international conference on Dublin core and metadata applications* (pp. 53–63). https://dcpapers.dublincore.org/pubs/article/view/3702

Lock, G. (2003). *Using computers in archaeology: Towards virtual pasts*. London: Routledge.

Lock, G., & Pouncett, J. (2017). Spatial thinking in archaeology: Is GIS the answer? *Journal of Archaeological Science, 84*, 129–135.

McCoy, M. D. (2017). Geospatial big data and archaeology: Prospects and problems too great to ignore. *Journal of Archaeological Science, 84*, 74–94.

McLaughlin, T. R., Whitehouse, N. J., Schulting, R. J., McClatchie, M., Barratt, P., & Bogaard, A. (2016). The changing face of Neolithic and Bronze Age Ireland: A big data approach to the settlement and burial records. *Journal of World Prehistory, 29*(2), 117–153.

Mills, B. J. (2017). Social network analysis in archaeology. *Annual Review of Anthropology, 46*, 379–397.

Niccolucci, F., D'Andrea, A., & Crescioli, M. (2001). Archaeological applications of fuzzy data bases. In Z. Stančič & T. Veljanovski (Eds.), *Computing archaeology for understanding the past* (pp. 107–116). Oxford: Archaeopress.

Niven, K. (Ed.). (2013). *Caring for Digital Data in Archaeology. ADS/digital antiquity guides to good practice*. Oxford: Oxbow.

Östborn, P., & Gerding, H. (2014). Network analysis of archaeological data: A systematic approach. *Journal of Archaeological Science, 46*, 75–88.

Plog, S. (1980). *Stylistic variation in prehistoric ceramics. Design analysis in the American southwest*. Cambridge: Cambridge University Press.

Richards, J. D. (1997). Preservation and re-use of digital data: The role of the Archaeology Data Service. *Antiquity, 71*, 1057–1059.

Robinson, I., Webber, J., & Effrem, E. (2015). *Graph databases: New opportunities for connected data*. Sebastopol: O'Reilly Media.

Ryan, N. (1992). Beyond the relational database: Managing the variety and complexity of archaeological data. In G. Lock & J. Moffett (Eds.), *Computer applications and quantitative methods in archaeology, 1991* (pp. 1–6). Oxford: Tempus Reparatum.

Schmidt, A. (2001). *Geophysical data in archaeology: A guide to good practice* (2nd ed.). Oxford: Oxbow Books.

Scianna, A., & Villa, B. (2011). GIS applications in archaeology. *Archeologia e Calcolatori, 22*, 337–363.

Sheehan, B. (2015). Comparing digital archaeological repositories: tDAR versus Open Context. *Behavioral & Social Sciences Librarian, 34*, 173–213.

Skibo, J. (1992). *Pottery function: A use-alteration perspective*. New York: Plenum.

Smith, R. (1973). *Pella of the Decapolis: The 1967 season of the College of Wooster expedition to Pella*. Wooster: College of Wooster.

Ullah, I. I. T. (2015). Integrating older survey data into modern research paradigms: Identifying and correcting spatial error. *Advances in Archaeological Practice, 3*(4), 331–350.

Watts, J. (2011). Building tDAR: Review, redaction, and ingest of two reports series. *Reports in Digital Archaeology, 1*, 1–15. Tempe: Arizona State University.

Weinberg, V. (1992). *Structured analysis*. New York: Yourdon.

Yoon, B. H., Kim, S. K., & Kim, S. Y. (2017). Use of graph database for the integration of heterogeneous biological data. *Genomics and Informatics, 15*(1), 19–27.

5 Visualizing Archaeological Data

Graphics reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations.

Tufte (1983: 13, italics in original)

Once the graph is understood, it can be used to make comparisons in a way that is nearly impossible from a table.

Gelman et al. (2002: 129)

Rather than being almost passive filler in an article, a "good" formal table becomes an active part of the study, able to bring central quantitative points to the reader's mind as quickly and nearly as succinctly as a well-designed graph.

Lewis (1986: 281)

"A picture's worth a thousand words." This cliché is only true if the picture clearly shows what its author intended, preferably, in the case of scientific pictures, without being misleading, intentionally or otherwise. However, in archaeology as in many other endeavors, the graphics that purport to describe and to some extent explain data are sometimes cluttered, confusing, or even misrepresentations.

Graphics allow us to display data visually. If used effectively, graphs can communicate complex information in ways that a viewer can interpret accurately. An inappropriate or poorly designed graph, by contrast, can be confusing or misleading. Among the many kinds of graphs that exist, it is important to select the right kind of graph for the kind of data you have and the point you are trying to make. Sadly, many of the computer packages that people use make it easy to generate inappropriate or misleading graphs. Remember that the sales and marketing people who are among the major users of such software have very different goals than archaeologists and communicating accurately and honestly is not necessarily one of them.

Among the questions you should ask yourself before you decide what kind of graph to use, or even whether to use a graph at all, are the following:

- What is the main point I want to convey?
- What scale or scales of measurement are in the data?
- Are the data continuous or discrete?
- How many dimensions should appear in a single graph?
- Will the graph show a frequency distribution?
- Who will use the graph? Is it for publication of final results or just for my own use?
- How will viewers want to use the graph? Will they want to extract detailed information or just compare information to get an impression of my results?

5.1 Tables

An alternative to a graph when you only need to show a small data set or when you anticipate that viewers may want to consult your raw data is a table. Tables are particularly effective when you only need to show a few values for a small number of attributes, on only two or three dimensions, or a frequency distribution with just a few categories. Other common uses are to show results of statistical tests (see Chap. 8) or radiocarbon data and calibrations (see Chap. 20). The rows and columns should be labelled in such a way that it is easy for viewers to interpret the table, without having to consult the caption, and including units in the column or row titles if there are any (not on individual values). You may draw attention to particular observations in the table, perhaps by bolding or color (Table 5.1).

Whenever the table displays either the number or the proportion of observations that fall into several categories, it is a simple kind of **frequency distribution**. Often the point of such a table is to show either that observations tend to

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_5

Table 5.1 Table showing the frequency distribution by Number of Identified Specimens (NISP) of the major animal genera from four Neolithic sites in Jordan, showing counts and (percentages), and extreme values highlighted by bolding

<table>
<tr><th rowspan="2">Site</th><th colspan="2">Ovis/Capra</th><th rowspan="2">Gazella sp</th><th rowspan="2">Sus sp</th><th rowspan="2">Bos sp</th><th rowspan="2">Equus sp</th><th rowspan="2">Total</th></tr>
<tr><th>Ovis sp</th><th>Capra sp</th></tr>
<tr><td rowspan="2">‘Ain Ghazal</td><td colspan="2">435 (70)</td><td rowspan="2">78 (12)</td><td rowspan="2">71 (11)</td><td rowspan="2">25 (4)</td><td rowspan="2">16 (2.6)</td><td rowspan="2">625</td></tr>
<tr><td>114 (18)</td><td>139 (22)</td></tr>
<tr><td rowspan="2">Wadi Shu‘eib</td><td colspan="2">83 (65)</td><td rowspan="2">6 (5)</td><td rowspan="2">24 (19)</td><td rowspan="2">10 (7.8)</td><td rowspan="2">5 (4)</td><td rowspan="2">128</td></tr>
<tr><td>6 (5)</td><td>30 (23)</td></tr>
<tr><td>Basta</td><td colspan="2">17,141 (86)</td><td>1769 (8.8)</td><td>40 (0.20)</td><td>649 (3.2)</td><td>431 (2.2)</td><td>20,030</td></tr>
<tr><td rowspan="2">Ba‘ja</td><td colspan="2">4793 (93)</td><td rowspan="2">181 (3.5)</td><td rowspan="2">8 (0.2)</td><td rowspan="2">59 (1.1)</td><td rowspan="2">102 (2.0)</td><td rowspan="2">5143</td></tr>
<tr><td>146 (2.8)</td><td>442 (8.6)</td></tr>
</table>

Note that faunal analysts cannot always distinguish *Ovis sp*. from *Capra sp*., and the report from Basta never did, so it is necessary to add an “Ovis/Capra” category. It is important to ensure that this *includes* the *Ovis sp.* and *Capra sp*. data, not just the bones that could not be assigned to a single genus, both for logical reasons and to ensure that the Ovis/Capra figures are comparable across sites (data from Makarewicz 2013: 242, but percentages rounded to two significant digits). This table has about the maximum amount of information you can expect viewers to absorb in a table

bunch up in one or very few categories, or to show similarities or differences between two or three frequency distributions. You can highlight some of these observations by color or bolding to make sure your viewer notices them.

Whenever the purpose of a table of frequency distributions is to reveal an association between two or three nominal or ordinal variables (the “dimensions”), it is called a **contingency table** (or cross-tabulation, e.g., Table 3.6). Ideally, such a table should allow viewers, at a glance, to identify what the association is and obtain at least a rough sense of how strong it is. In a 2 x 2 contingency table, distributions with strong associations between the two dimensions have large numbers in the two cells in one diagonal, and low numbers on the other diagonal. But there are statistical methods that you can use to assess such associations quantitatively, such as Pearson’s chi-square test, and you can measure the strength of the association with the φ-coefficient or Cramér’s V (Baxter 2003: 128–131; Drennan 2010: 182–190; Shennan 1997: 104–118).

One thing you should not do, unless the table is just for archiving information (e.g., a reference table, appendix or supplement), is to present a huge table with a sea of numbers that no one can interpret without considerable time and effort. Large tables of data are poorly suited for disseminating complex information clearly or quickly, making them particularly ineffective in a lecture or conference presentation. In such cases, a well-planned graph is usually preferable. In other cases, it might be possible to break the large table down into smaller tables that highlight the points you would like to make.

Ehrenberg (1981) suggests some guidelines for more effective tables that include:

- Limit significant digits to two
- Organize the table’s layout so that it guides the viewers’ attention to the most important information or facilitates key comparisons
- Provide marginal averages
- Order the rows or columns by their marginal averages (or some other size measure), but standardize the order of multiple tables with similar data, with larger values in the top rows, if possible
- Organize the table in such a way that the numbers people will most want to compare are in the same column, instead of the same row
- Use the table caption to summarize the table briefly and point out the main patterns or anomalies

As an exception to some of these guidelines, one could point out that some archaeological patterns are chronological, in which case it makes sense to order the rows stratigraphically or chronologically, rather than by marginal averages, and with the earliest data at the bottom of the table to mimic stratigraphic order. Also, one could use marginal medians instead of means.

For more advanced readers, Lewis (1986) provides a really good summary of the effective use and analysis of contingency tables, including Tukey’s (1977) procedure of “median polish” to transform the figures in a table in such a way as to highlight meaningful structure, and use of log-linear models to identify the interactions among variables in multi-dimensional tables (see also Baxter 2003: 131–136).

5.2 Common Varieties of Archaeological Graphs

What follows are descriptions of the features and potential uses of the kinds of graph most often used in archaeology and some that could be useful but currently are not common. A few are best used only in exploratory analysis, while others are suitable for publication and presentations. In every case, it is important to keep in mind the scales of measurement involved in the data, whether the data are one- or multi-dimensional and discrete or continuous, and whether your intention is to show a distribution or make a comparison.

5.2.1 Box-and-Whisker Plots and Stem-and-Leaf Plots

These plots can be useful to compare batches of data, quickly and without losing detail, during "exploratory data analysis" (Tukey 1977) while you hone a research design. Both are used for interval-scale frequency distributions on a single dimension (Drennan 2010: 4–11).

The stem-and-leaf plot is appropriate for a small amount of interval data that you want to tally manually (usually with pencil and paper) without losing any of the original measurements. It can be a prelude to making a bar graph or histogram, especially if you draw it manually. It begins by drawing a scale ranging from just below the lowest measured value to just above the highest one, the scale consisting of a list of values in increments, typically of ten or five units each, and omitting the last significant digit. You then build up the plot by recording the last significant digit of each measurement next to its appropriate interval. For example, in Fig. 5.1, there are two artifacts with measured lengths of 149 mm; those are each indicated by a "9" next to the "14" on the plot. There is also an artifact measuring 128 mm, represented by an "8" value next to the "12". Stem-and-leaf plots have diminished importance today because of the prevalence of digital tools for data analysis and presentation, and they are not appropriate for publication or presentations.

A box-and-whisker plot summarizes one-dimensional data more completely, so you lose most of the individual measured values but retain a good impression of how the data are distributed. This plot shows a wide box for the interquartile range (IQR, middle 50% of the values), a perpendicular line segment marking the median, and thin bars or "whiskers" extending away from the box to show the maximum and minimum values (i.e. range). Some versions use the closely related Median Absolute Deviation instead of IQR, and it is also common to identify outliers separately by dots or small circles. The only measured values that you can extract from the graph are the maximum and minimum values. This kind of plot is useful in exploratory analysis and sometimes features in archaeological presentations (Fig. 5.2).

Fig. 5.1 Stem-and-leaf plot (left) and a corresponding tally (right) to show two different ways to record the frequency distribution of blade lengths in millimeters. The former includes all the raw data, including the last digit, while the latter, like a histogram, only tells you how many observations fell within each interval

20	7	I
20		
19		
19	3	I
18		
18	1	I
17	6	I
17	1	I
16		
16		
15	9	I
15	46	II
14	997	III
14	21	II
13	8	I
13	43	II
12	8	I
12	123	III
11	67	II
11	3	I
10	8	I
10		
9		
9	2	I
8	8	I
8		

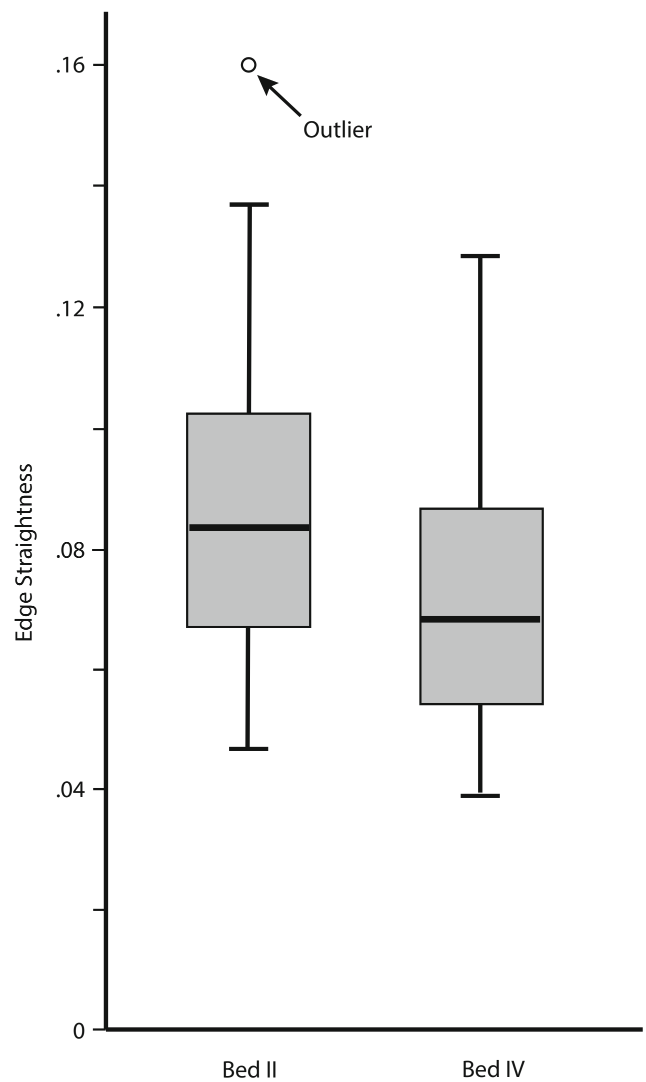


Fig. 5.2 Box-plot of edge straightness of bifaces from Olduvai Gorge Beds II and IV. (After Shipton 2018: 117)

5.2.2 Azimuth Plots

Another kind of plot that, like the stem-and-leaf plot, does not lose any of the original data is one designed to show distributions of orientations (Fig. 5.3). It shows a circle, with North at the top, and radiating line segments to indicate every orientation in a sample. It can show the orientations of houses, temples or chamber tombs (e.g., Prendergast 2016), astronomical alignments in a monument, the head positions of skeletons in a cemetery, or the aspect of sites on slopes.

5.2.3 Bar Graphs

Archaeologists sometimes confuse bar graphs, which are intended for frequency distributions of discrete (usually nominal) data, with histograms, which are frequency distributions for interval or ratio-scale (and often continuous) data (Drennan 2010: 71–73). Bar graphs can be effective replacements for small tables of frequencies because they make it easy to visualize how the numbers or proportions of observations in a number of categories vary. Most computer spreadsheets and statistics packages, as well as dedicated graphics software, can generate a bewildering array of bar graphs, often incorrectly labelled as histograms. You should avoid the temptation to use most of these, which are often overly complicated, have too many embellishments that could distract your viewers, or, worse yet, show a separate bar for every observation instead of making a frequency distribution.

A bar graph simply displays, by the heights of the bars, how the number of cases varies by category. It is thus a type of frequency distribution. The bars are separated from one another to signal that the values on the x-axis are discrete categories rather than continuous values. The heights of the bars are proportional to both the number and the proportion

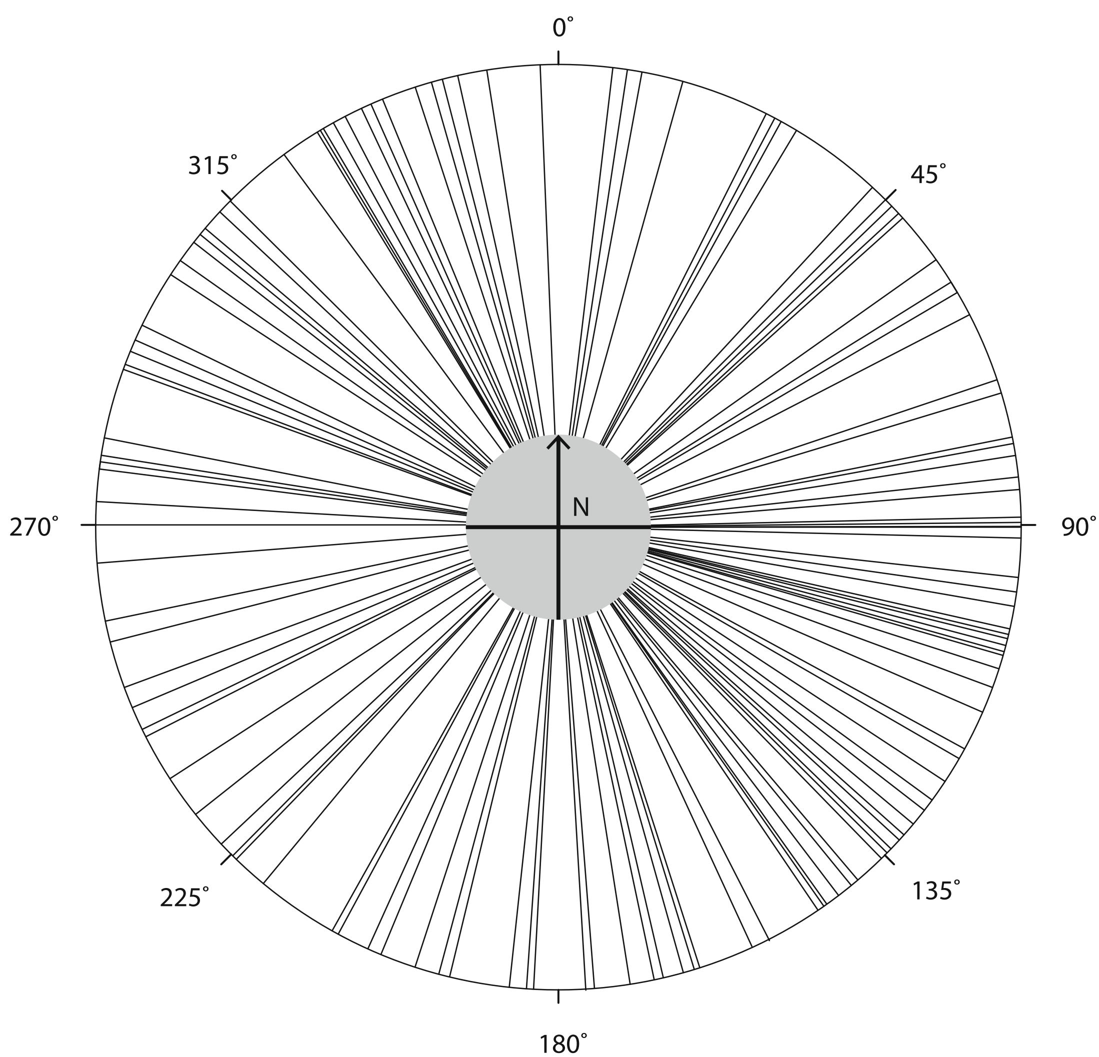


Fig. 5.3 Azimuth plot of orientations in 132 Irish “court cairn” tombs. (Y. Salama, After De Valera 1959: plate 35)

of observations for each category. At least one scale on the y-axis allows viewers to measure the heights of the bars to determine the number or proportion that each bar represents, but the most important feature of this kind of graph is that, even without consulting that scale, it is easy to see at a glance whether any category stands out for either a high or a low number of observations. It is not uncommon to show two y-axes, one for number and the other for proportion of observations, but to avoid clutter you should only do this if it has a purpose.

The y-axis is usually on a linear scale, but could be logarithmic, square-root, or something else. A non-linear scale can be helpful if you need to show both very small and very large values on the same graph. However, if the scale is non-linear, you should make that very clear in the label and title or caption so as not to mislead viewers (see "graphical integrity" below).

Sometimes it is tempting to put too many kinds of observations on a single bar graph. This can confuse viewers and make it more difficult to compare data in meaningful ways. It is usually much better to show only one dimension on each bar graph, although you could "stack" a series of bar graphs at the same scale and with the same categories if you want to make comparisons, for example, among sites. This is generally a much better solution than putting many bars with different colors or hatching next to each other on the same bar graph (see Fig. 5.4), although groups of very few bars may be acceptable. If you have several groups of data to compare, divide them into meaningful sub-graphs and arrange them in ways that facilitate comparison, rather than shuffling all the data into a single, confusing graph. Another solution, if you want to compare two data sets, is to rotate the bar graphs and reflect one of them, so that they share an x-axis (Fig. 5.5).

A common archaeological version of the bar graph is the "battleship curve" or "spindle graph" used in seriation (fig. 5.6a). Ford (1962) made this popular with archaeologists, but something very similar has a longer history (Olson 1930; Sayles 1937; Lyman et al. 1998). In Ford's version, it is

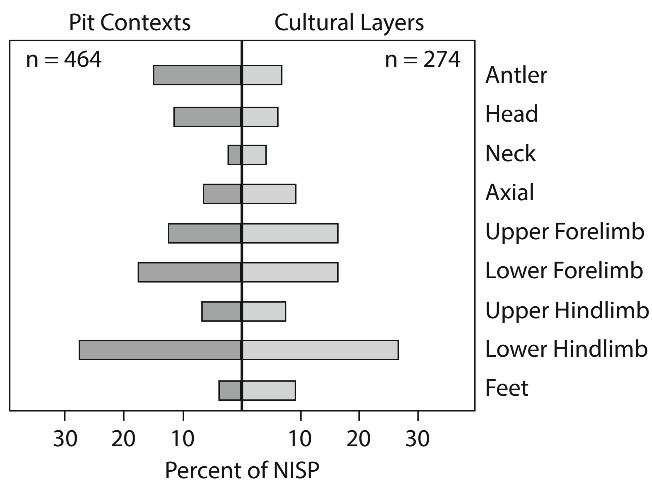


Fig. 5.5 Two back-to-back bar graphs to facilitate comparison of distributions of body parts by NISP among deer remains in pits and other cultural layers at the site of Gomolava, Serbia. (Data from Orton 2012: 334)

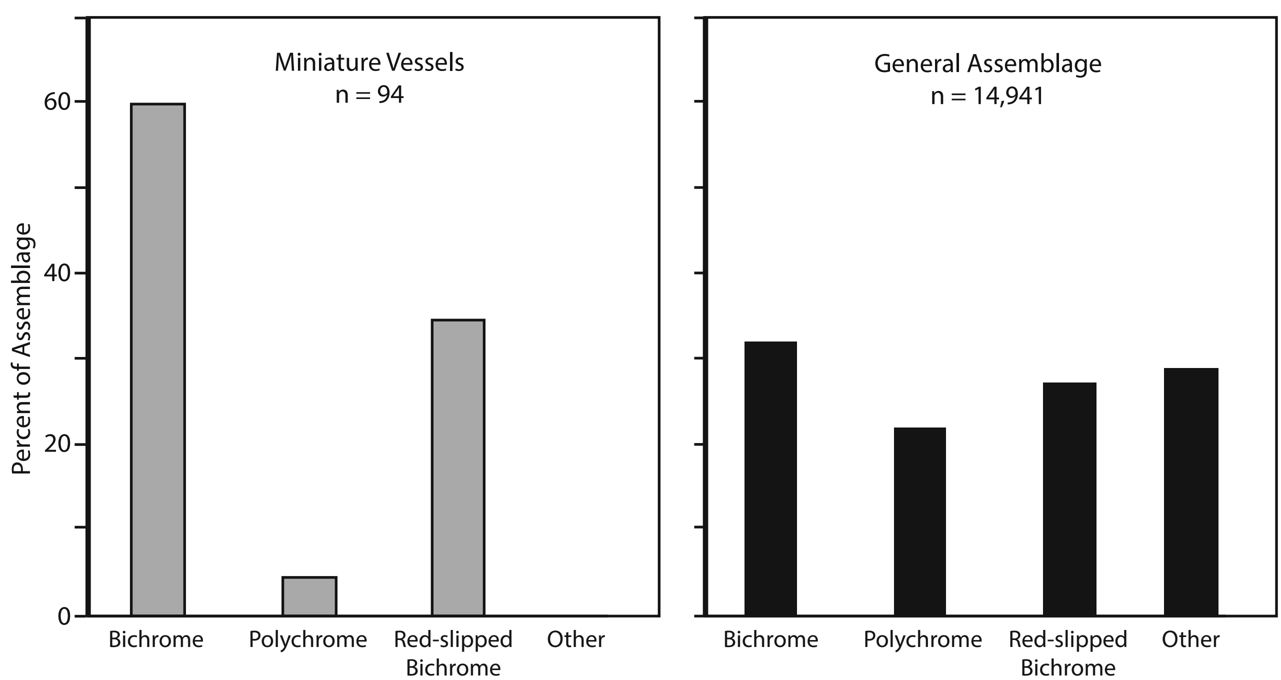


Fig. 5.4 Bar graphs to show the distribution of surface treatments among miniature vessels and the general assemblage at the Homol'ovi I site in northeastern Arizona. (After Fladd and Barker 2019: 114)

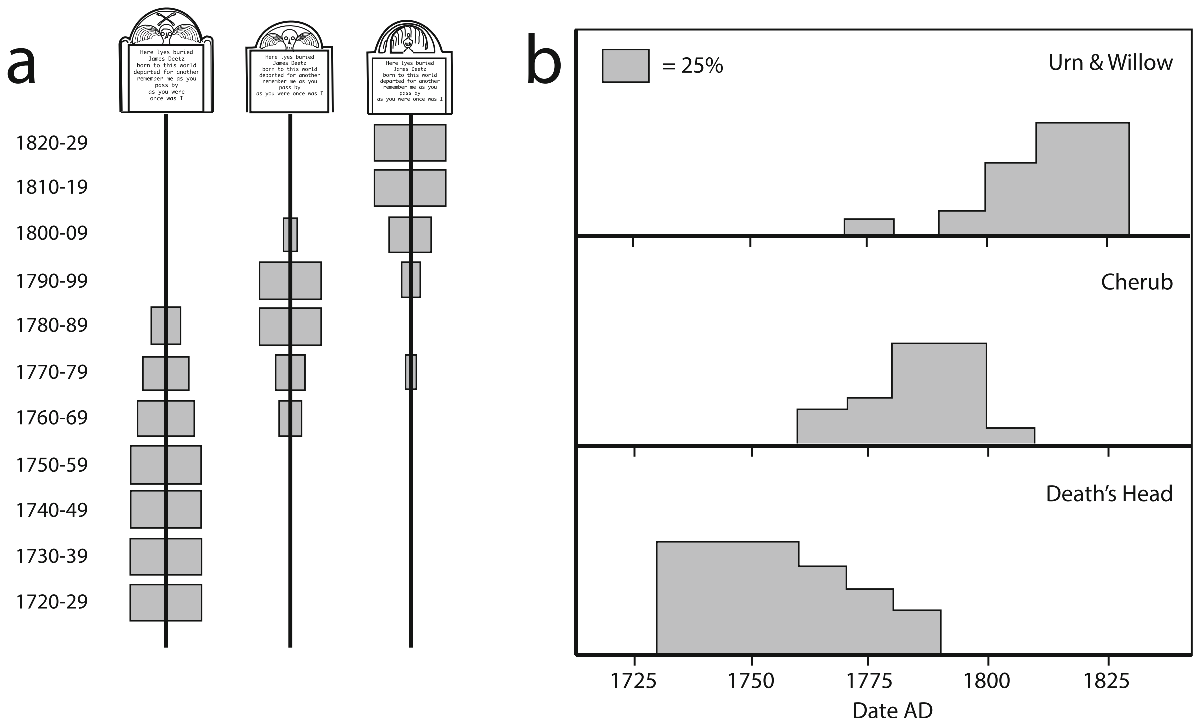


Fig. 5.6 "Battleship curve" (**a**) of tombstones from New England. (After Deetz 1967) and (**b**) the same data shown as histograms

really a series of bar graphs rotated 90° and reflected about the x-axis so that the bars are centered. It contains no more information than a regular bar graph (fig. 5.6b) but some would argue that it allows more efficient use of space by allowing distributions to "nest" closely together. The best thing about these graphs is the rotation; it allows us to arrange the bars in such a way that the oldest category is at the bottom and youngest at the top, a useful graphic metaphor for the passage of time.

5.2.4 Stacked Bar Graphs

These are much like standard bar graphs except that the bar heights are proportions and the bars are stacked so that they add up to 1.0 or 100%. Varying the color or hatching of the bars allows viewers, through a key, to distinguish the categories in segments of each bar.

These can be somewhat effective when you want to compare something like the distribution of animal taxa or lithic raw materials among several sites or contexts within sites, when the number of taxa is relatively small, and when the number of sites or contexts is high enough that a solution like putting regular bar graphs side-by-side might be cluttered and hard to follow (Fig. 5.7). However, it is important to think carefully about what order to place the taxa in the bars, and to keep it consistent for all bars. Usually you should make it easy for viewers to compare the most abundant taxa, in which case they should be at the bottom or top of each bar, but sometimes there will be some less numerous taxon that you want viewers to notice, perhaps by its color.

However, stacked bar graphs tend to be hard to read, and there are many cases where the data would be better served with a group of ordinary bar graphs (Lyman & Faith 2018: 442) that make it easier for viewers to compare the heights of bars across categories or contexts. When there are three categories, a ternary graph (pp. 77 and 297) might be a better choice.

Incidentally, stacked bar graphs with only two categories in each bar simply waste space. If you only have two taxa, such as domesticated and non-domesticated animals, or presence and absence, just pick one and use a regular bar graph for the data. Do not insult your viewers' intelligence by stacking 40% on top of 60% or "absent" on top of "present."

5.2.5 Time-Series Bar Graphs

Because of their interest in change over time, archaeologists have sometimes adopted a type of time-series graph that represents the presence of some feature or artifact type or the discontinuous occupation of sites by bars along a time scale.

This is a convenient way to see such things as which sites were simultaneously occupied, where settlement discontinuities occur, or the associations of artifact types

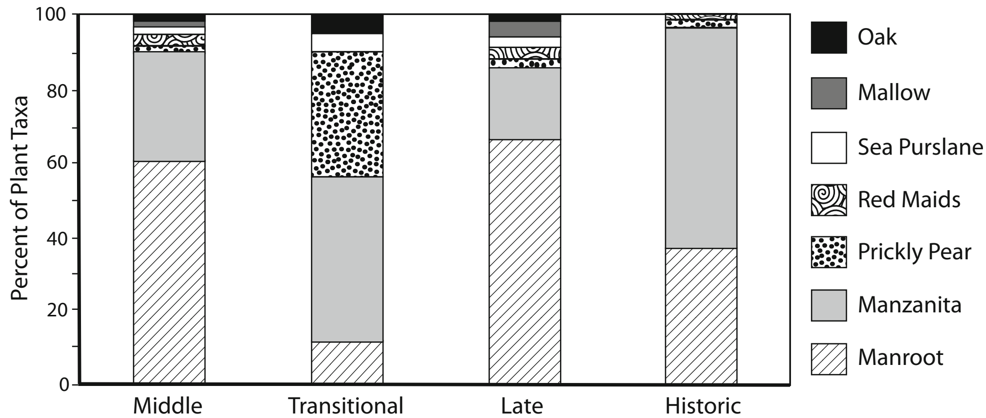


Fig. 5.7 Stacked bar graph of carbonized plant taxa by time period at Santa Cruz Island (after Arnold & Martin 2014: 241). Note the distracting moiré effects (patterns not used in the original), and that these data could have been shown effectively in a time series instead

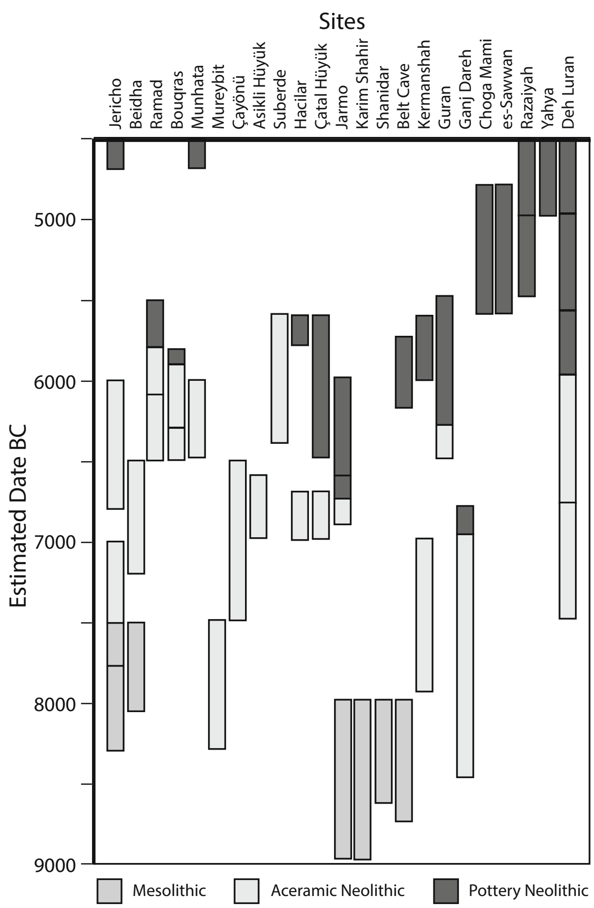


Fig. 5.8 Time-series bar graph indicating the times (as once understood) when each of several Near Eastern sites was occupied. Gaps in the bars indicate site abandonments (modified from Singh 1974). Note the time arrangement with oldest at the bottom, as in stratigraphic order

during certain periods (Fig. 5.8). In these graphs, time should "flow" either from left to right or bottom to top, the latter reflecting stratigraphic order. The data in each column of such a graph is usually at a dichotomous scale (presence or absence), while the time axis is an interval or, if necessary, an ordinal scale. Unlike normal bar graphs, these are not frequency distributions.

5.2.6 Pie Charts

These are very common in the popular media to show proportions of things in a frequency distribution and are now very easy, perhaps too easy, to generate with computer software. They consist of a circle subdivided into segments by lines that radiate from the centre, with the angles between radii proportional to the proportion of observations in each category. While the media tend to use this for things like "how your tax dollar is spent," archaeologists have often used it to display things like faunal distributions (Fig. 5.9).

Because they are for showing frequency distributions on a nominal scale, one could just as easily use a bar graph to show the same data and, in fact, bar graphs are usually superior in several respects. First, although the area of each "pie slice" is proportional to the number of observations, humans find it easier to estimate areas of rectangles than areas of wedges, and easier to compare heights than angles (Tufte 1983: 55). Thus, bar graphs are easier to interpret and compare. Many users of pie charts try to compensate for this by putting labels on each wedge to indicate the values of the proportions. However, this makes the graph redundant, and you might as well use a table. In many cases, either a bar graph or a small table works better than a pie chart.

An exception is when you want to show how some frequency distribution varies across points or spaces on a map. You can place a small pie chart on each location and even vary the charts' sizes so that their areas are proportional to sample size, while the slices of each circle, as usual, indicate the proportion of each type of artifact, plant, animal, raw material, etc. (Fig. 5.10).

5.2.7 Windflowers

Windflowers are named after their common use to indicate wind directions. Like azimuth plots, they can be useful whenever it is helpful to show orientation or directions. Like bar graphs, however, they summarize the data more by using bars that radiate from the center with a height proportional to some value, such as the number of days or hours that the wind blew from a particular direction. Although you could show exactly the same kind of information in a regular bar graph, the attraction of this kind of plot is that it conveys the actual directions that are of interest quite well (Fig. 5.11).

Archaeologists have used windflowers or similar types of graphs called rose plots (Fig. 5.12) to summarize the orientations of houses, temples or chamber tombs, astronomical alignments, the head position in skeletons (Pankowská and Monik 2017), or the aspect of sites on slopes. Palaeolithic archaeologists have also used them to summarize the dip angles (or inclinations) of flakes and blades when excavated. One of the disadvantages of rose plots over windflowers is that they create distortion by exaggerating large values by violating the principle of proportionality by area (see "General Principles," below).

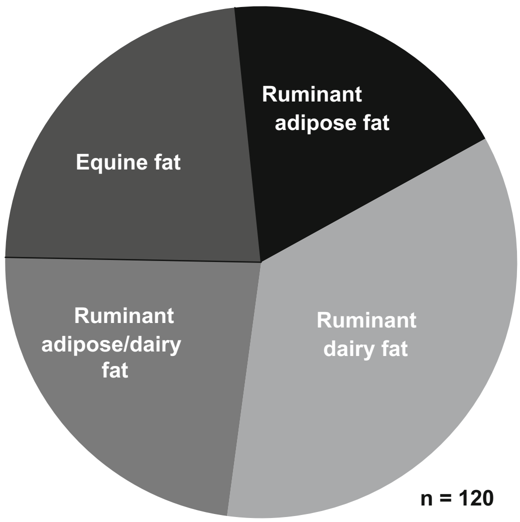


Fig. 5.9 Pie chart of the distribution of four different sources of fat following analysis of chemical residues in pottery. (Y. Salama)

5.2.8 Histograms

Histograms are appropriate for displaying frequency distributions of data in a single dimension on an interval or ratio scale (Drennan 2010: 11–15). Although they may look

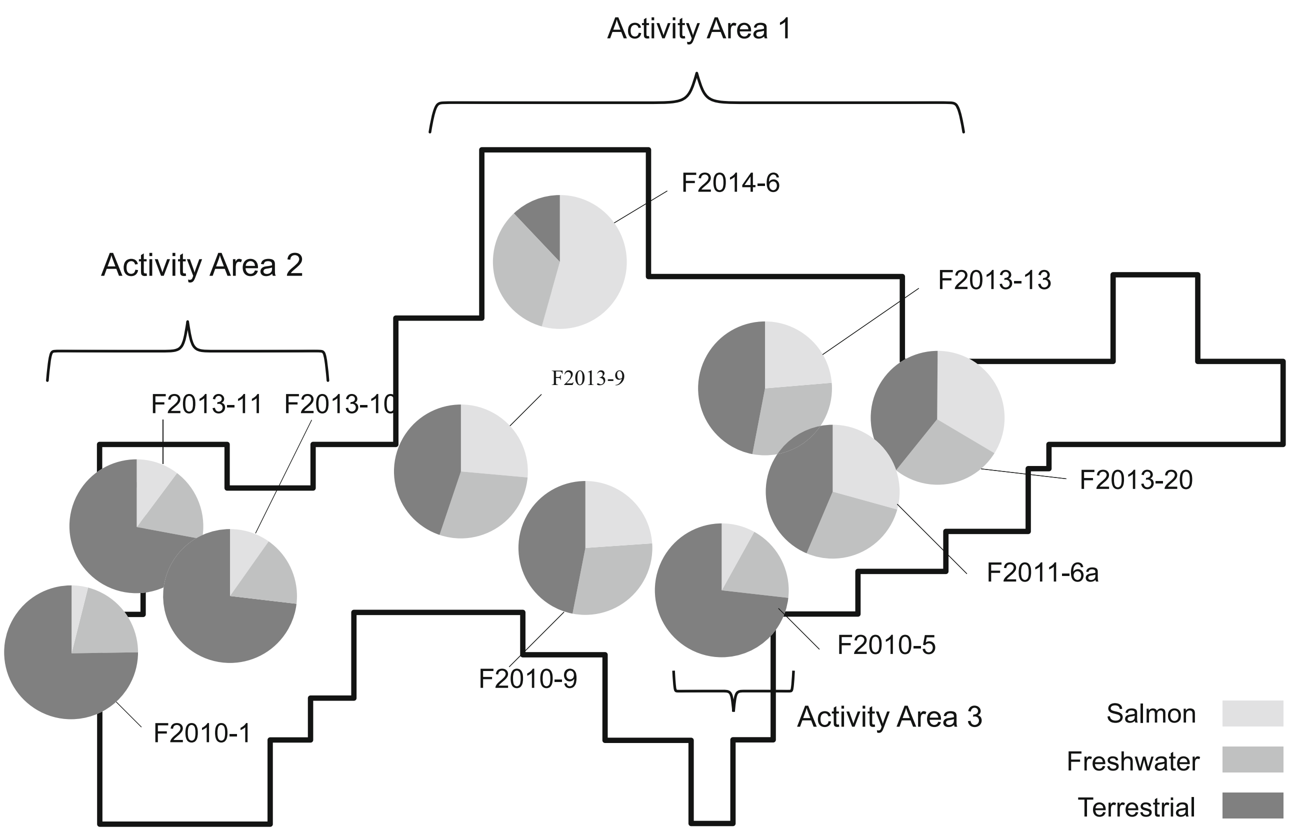


Fig. 5.10 Map with pie charts to show the proportional variation in food groups across space in the Upward Sun River site in central Alaska. (Y. Salama, after Choy et al. 2016: 9760)

much like bar graphs, they are actually quite different. One difference is that, for continuous data, it is the area, not the height, of bars that is important. This allows us to use bars of varying width without distorting the data. Another difference is that we can signal continuous data by bars that are contiguous, while separating the bars indicates discrete data (Figs. 5.13 and 5.14).

In a histogram, the *x*-axis is an interval or ratio scale. If that scale is also continuous, it is necessary to group data into intervals and count how many observations fall into each interval. If you are doing this manually, and your histogram is for continuous data, the procedure is as follows:

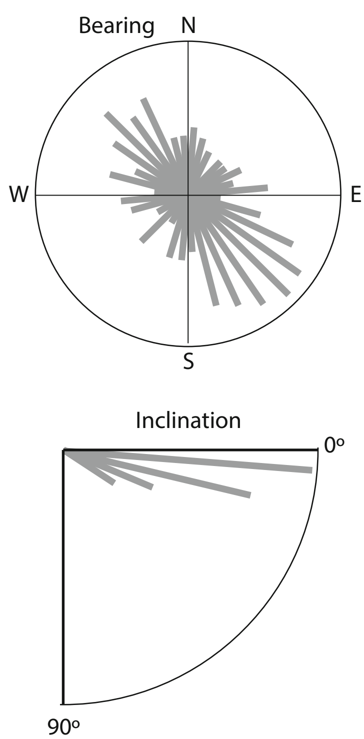
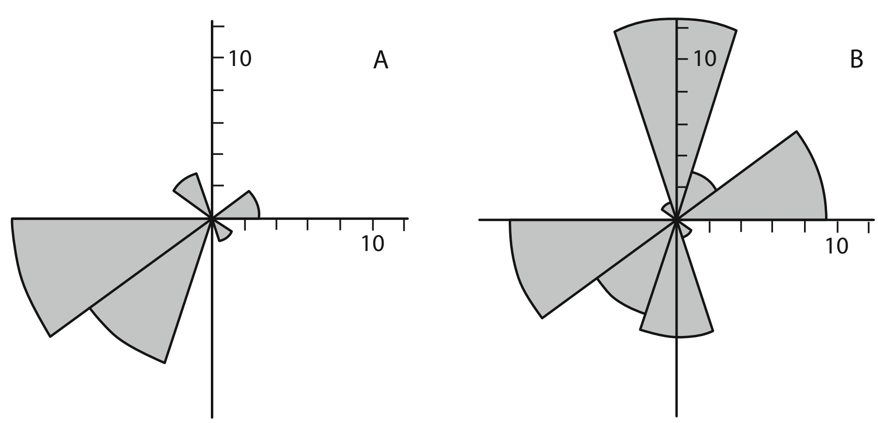


Fig. 5.11 Windflowers to show the distribution of horizontal orientations (compass bearings) and dip angles (inclination) of artifact refits in layer 2 at La Ferrassie, France. (Data from McPherron 2018: 12)

- Select an interval for your “bins,” assuming for the moment that you will use equal intervals. You will want an interval narrow enough to show variation in the data but large enough to yield peaks and valleys. You may need to try several different intervals before you find one that presents your data effectively, or you might select an interval that divides your data into eight or ten groups. The intervals do not have to be divided at 0, 5, 10, etc., but should be based on less arbitrary criteria.
- Uniquely define the groups, not with 0–4, 4–8, 8–12, for example, but 0–3.99, 4–7.99, etc., depending on how many significant digits your data show. However, you should not label your intervals that way (just label some of the boundaries between intervals). Also, do not use an interval like “>10,” as it has an infinite width.
- Tally the number of observations in each interval.
- Draw the histogram so that the area of each bar is proportional to the number of observations in each interval. The easiest way to do this, if the intervals are equal, is to measure their heights. As noted below, however, equal widths are not always the best solution, or even possible. To signal that the data are continuous, ensure that there are no gaps between bars (or even lines) except when there is an empty interval (no observations).
- Draw a rectangle somewhere in a blank part of the graph to show how much area on the graph corresponds to one observation, or ten, or 100. This provides a scale for your bars and makes it unnecessary to show a vertical or y-axis. Should you decide to do so, however, make sure that you label the units on the y-axis as “number of observations per interval *x*” (filling in 5 mm or whatever the correct interval is for *x*).
- As with all graphs, make sure that the scale is labelled, including the units, but do not over-label it. Give the graph a title or caption.

For histograms with discrete data, such as one showing the number of hearths per Iroquoian longhouse, we make the histogram with the following differences (Fig. 5.14):

Fig. 5.12 Rose plots to compare the angular distributions of body orientations in Early Bronze Age Grave Burials (**A**) and Storage Pit Burials (**B**) in Moravia. (Y. Salama, after Pankowská & Monik 2017: 922)



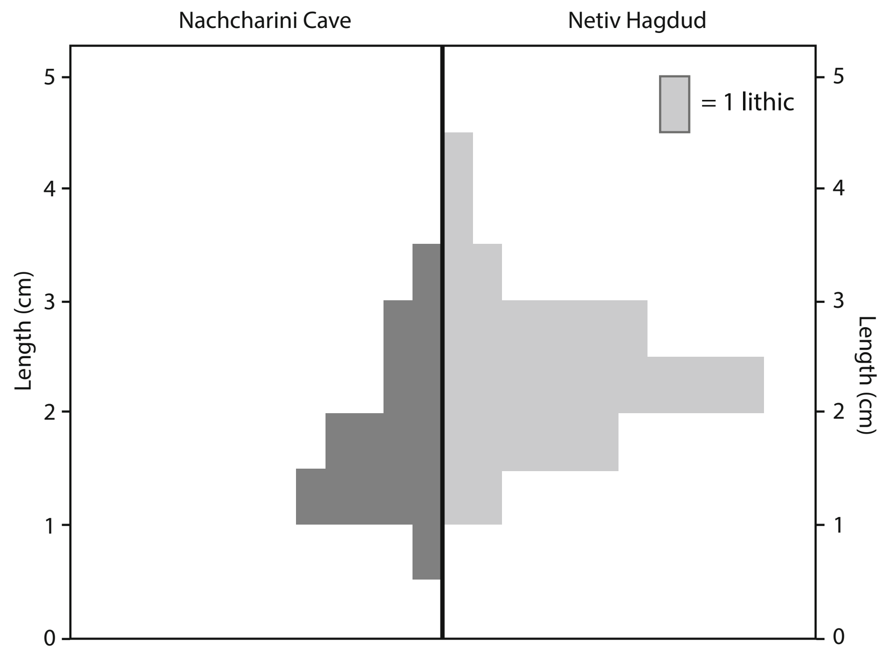


Fig. 5.13 Two histograms with continuous data, shown back-to-back to facilitate comparison of distributions of lengths of projectile points at two early Neolithic sites in Lebanon and Israel. (After Rhodes et al. 2020)

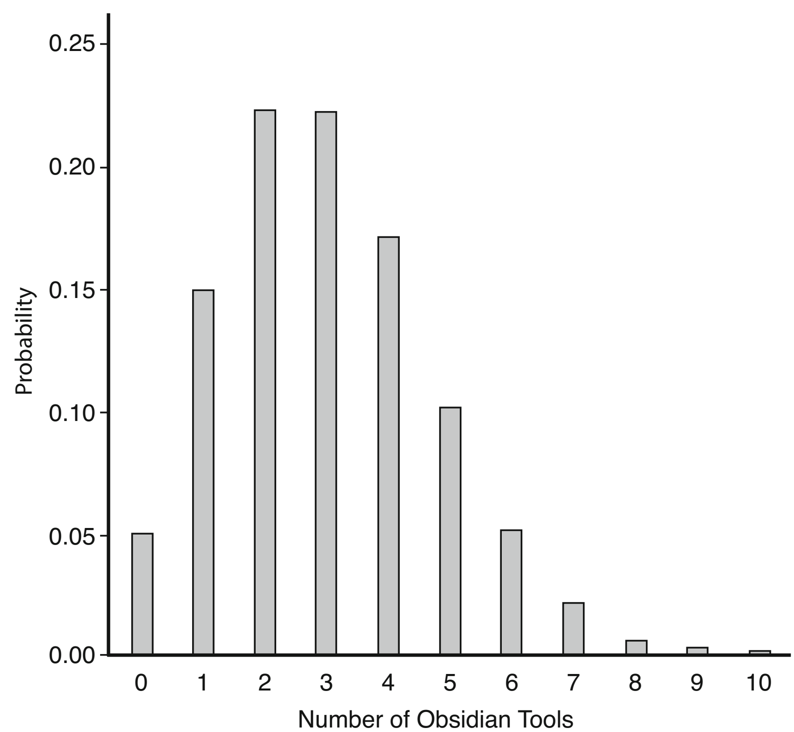


Fig. 5.14 Histogram with discrete data, in this case, a Poisson distribution with a mean of 4 (see Chap. 8)

- The interval is set by the use of a discrete (usually positive integer) scale, so 0, 1, 2, 3, or 4 hearths per house (never 3.7 hearths).
- We use separate, line segments or narrow bars centered on the values they represent.
- We use a labelled y-axis and the heights of the bars to indicate how many cases take each value (here areas are not relevant, as lines have no width and the interval is constant).

Continuous histograms with unequal intervals: Often, it is desirable or necessary to vary the intervals in a histogram. For example, the interval for a histogram of particle sizes may be determined by the mesh sizes of screens, or one intended to show variation in numbers of sites by time period would usually have intervals that correspond with time periods of various duration, rather than something like centuries. If we were to construct histograms with height representing frequency, and ignore the fact that the intervals are unequal, the result would give undue weight to the broader intervals. In other words, as our eye measures the "importance" of each bar by the area it takes up, it would exaggerate the contributions of the larger intervals. Similarly, if you have unequal time intervals on your x-axis, but display them as though they are equal, you will exaggerate the importance of the shorter time periods. This is a very common mistake among archaeologists, sometimes leading them down the path of completely incorrect conclusions (Fig. 5.15a).

The way to avoid such a mistake is to ensure that frequencies are proportional to area. So, if two intervals have the same number of observations but one interval is twice the width of the other, the wide bar should be half the height of the narrower bar (Fig. 5.15c). This approach is usually good when most of the intervals are equal, with only one or two exceptions. Another way to do it is to make sure that your bar heights are accurate when measured on a scale of "number of observations per interval *x*". Furthermore, if the x-axis is a time scale, and the intervals are

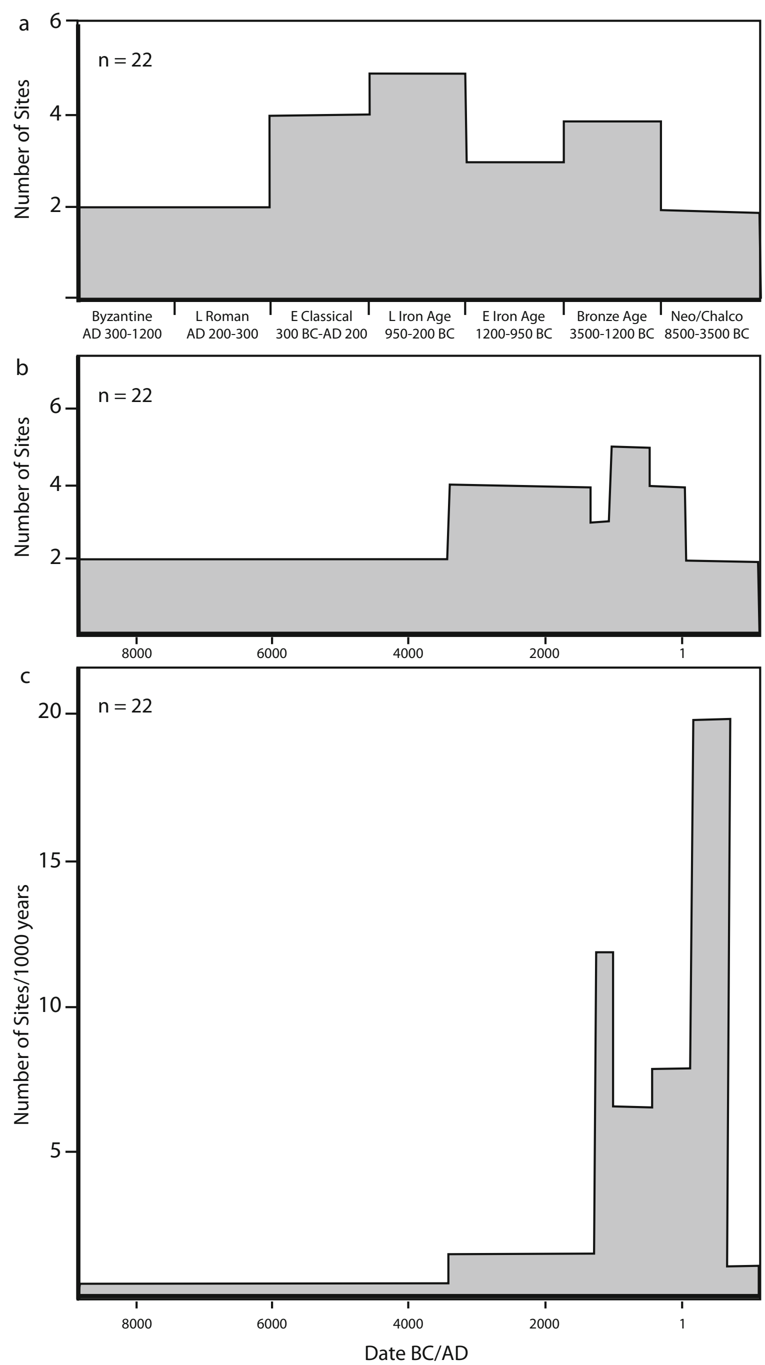


Fig. 5.15 The histogram in (**a**) distorts the information by treating unequal intervals as though they are equal (and also has time in reverse order), that in (**b**) corrects for the period durations but, unless all the sites were continuously occupied throughout their period, distorts histogram shape by using heights, and (**c**) corrects this effect by making the areas in each interval proportional to frequency

periods of unequal duration, try to show those intervals as accurately as you can (usually there will be error in estimated duration, but it may be relatively small) and, if possible, label the scale with a time axis in years, even if you have to estimate it with, for example, radiocarbon dates.

Before leaving histograms, it is necessary to point out that their overall shape is extremely sensitive to even small changes in the position or widths of intervals (Fig. 5.16). Consequently, you should not make too much of histogram shape unless it holds up to "tweaking" of the intervals, and should be especially wary if the histogram is based on a very small sample or if the intervals are very small or very large compared with the dispersion in the data. Where you need to be confident of the shape of a frequency distribution, it is better to use an ogive (cumulative frequency distribution), which does not have this disadvantage.

It is necessary to note that many of the "histogram" functions in standard software do not create histograms at all, and you should avoid them at all costs. Rather than grouping continuous data into intervals and tallying them, these just show a separate bar for every observation, and typically present them in an arbitrary order or alphabetically by the observation labels. If you must use this kind of software, you need to trick it by sorting your data and using subtotals for each interval you want to use as the data for your graph.

5.2.9 Broken-Line Graphs and Time-Series

The place of line graphs is to show variation in some value over a continuous scale that is often time, in which case we call it a time-series. While they may appear somewhat similar to histograms, the difference is that they are not frequency distributions; the value on the y-axis is not a count or proportion of observations per interval, but rather the value of some attribute or statistic, such as temperature, proportion of decorated pottery, mean length, or median house-pit area. The broken line on the graph consists of line segments joining a series of measured points, taken at regular or irregular intervals. To emphasize that these are not frequency distributions, you should avoid shading the area under the broken line.

One of the attractive features of broken-line graphs is that we can place two or three broken lines on the same graph without sacrificing clarity or making it too cluttered. This

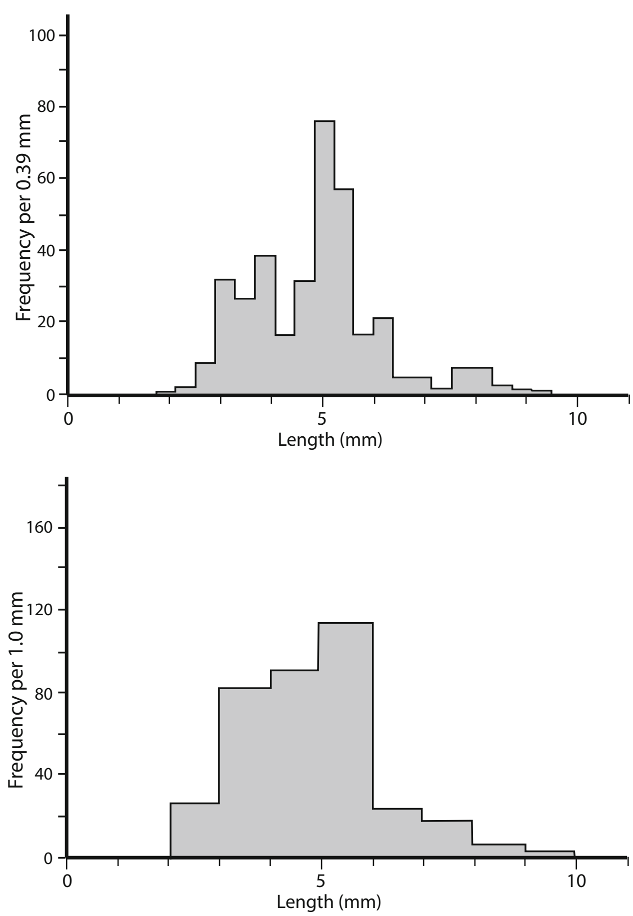


Fig. 5.16 Effect of "bin" size and location on histogram shape. All four histograms show the same data, but with different "bin" boundaries

makes comparing them really easy. Another is that we can put error bars on the points to indicate how much statistical or measurement error each one entails (Figs. 5.17, 5.18).

A time series is a broken-line graph in which time constitutes the scale of the x-axis. As archaeologists are typically interested in change over time, this is a particularly important class of graph. If done properly, sudden periods of change appear as abrupt change in slope (Fig. 5.18). However, if the time scale is distorted, for example, by showing time periods of equal width when in fact they vary substantially in duration, this may provide a highly misleading impression of rates of change (Fig. 5.19). As much as possible, archaeologists should attempt to depict time intervals accurately in these graphs, showing error bars, if necessary, to indicate uncertainties in timing.

Broken-line graphs are usually depicted horizontally but, as with seriation graphs, there are sometimes good reasons to rotate them 90°, especially when the x-axis represents time. To make time series easily interpretable, it is best to have time "flow" either from left to right or from bottom to top, the latter being a metaphor for stratigraphic order (Fig. 5.20).

A variation on the time-series graph that is common in fields like palynology is actually a type of frequency distribution that uses bars or filled broken lines to represent, for example, the abundance of pollen taxa in increments of a core through deep deposits, with depth as a proxy for time (see p. 350).

5.2.10 Cumulative Frequency Graphs (Ogives)

Ogives are particularly useful when you or a viewer might want to know what proportion of the data lie above or below some value, or between two values. It is also valuable for determining whether a frequency distribution is even or

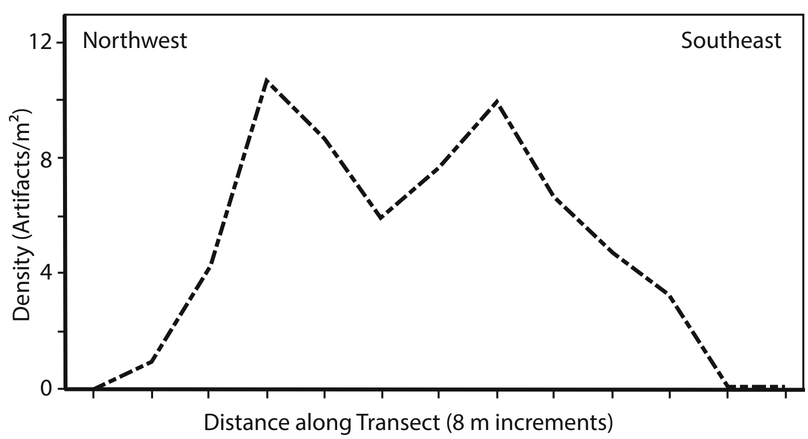


Fig. 5.17 Broken-line graph showing interpolated variation in artifact density along a transect across an artifact scatter in Lefkas, Greece. (Data from Gallant 1986: 407)

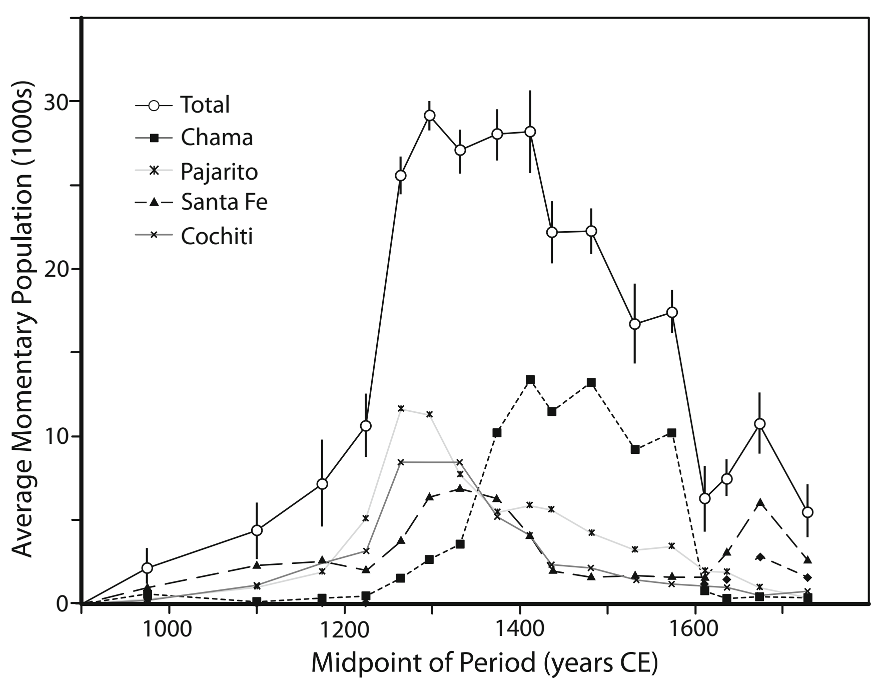


Fig. 5.18 Time-series graph showing how estimated momentary population sizes changed over time (after Ortman 2016). Steep slopes on the broken lines indicate rapid population growth or decline and bars on total population represent standard errors (here conjectural)

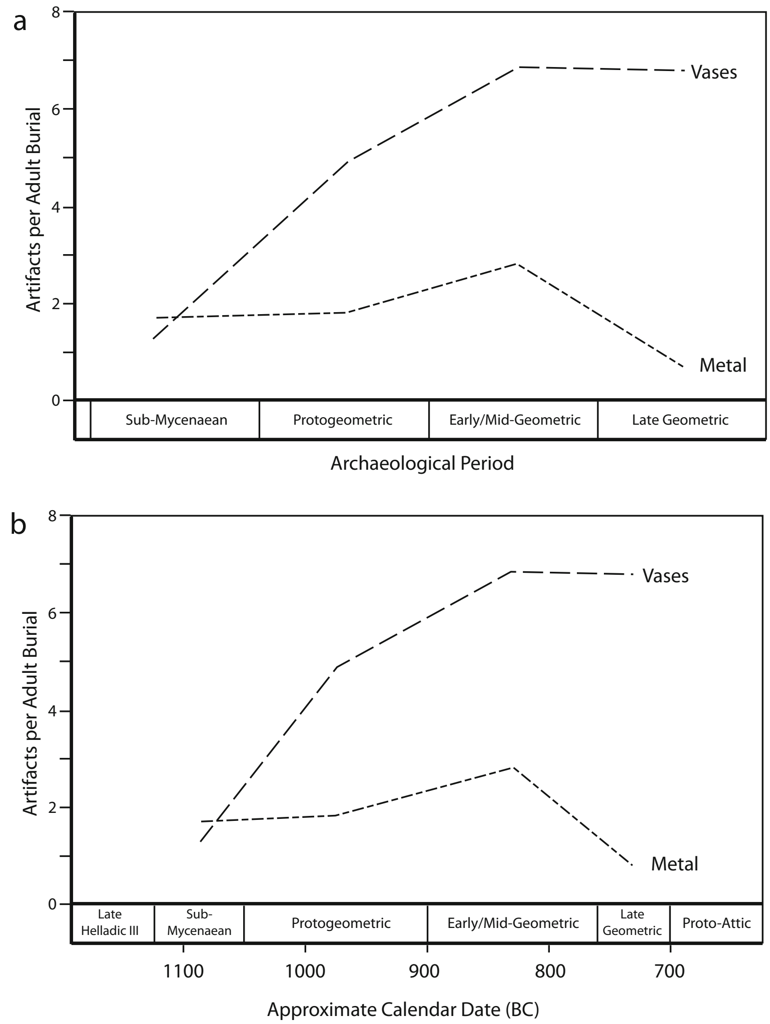


Fig. 5.19 Misleading time-series with unequal periods shown as though they were of equal duration (**a**) and a corrected version (**b**) with the approximate duration of each period indicated in years and points at the midpoints of periods (data from Morris 1987: 141). The latter more accurately represents rates of change

uneven, and for comparing two or three frequency distributions on interval or ratio scales (Orton 1980: 171).

Although it is possible to use cumulative frequency graphs for ordinal scales, it is most often used for continuous interval or ratio data (x-axis) and a measure of relative abundance (proportions or percentages, y-axis). As we pass from left to right along the x-axis, the values on the y-axis accumulate, so that they always rise from 0 to 1.0 or 100%. Making an ogive involves much the same steps as a histogram except that the frequencies are converted into proportions and the value assigned to each interval is summed with all the intervals to its left before being represented as a point. Once we join the points plotted in this way, the broken line rises from the bottom left corner of the graph to the upper right corner (Fig. 5.21).

One of the important applications of this type of graph is in a statistical test, the Kolmogorov-Smirnov test (Shennan 1997: 53–61), which is useful for comparing two distributions on the ordinal, interval or ratio scale. When you plot the two cumulative distributions, the maximum vertical difference between the two broken lines is a good measure of how different the two samples are ("D" in Fig. 5.21).

A common use of these graphs is to make it easy for viewers to see what proportion of a distribution is greater or less than some value of interest. For example, we might want to know what proportion of houses had areas greater than 30 m^2, or what proportion of graves had less than three grave goods. In non-archaeological applications, social scientists often use ogives to analyze income distributions by what they call the Lorenz curve (Gastwirth 1972). A straight, diagonal line indicates even variation in income (equal numbers of poor, middle-income and rich individuals or families), a highly concave distribution indicates that there were many poor families and most of the income went to very few people (e.g., the "top 1%"), and a slightly convex distribution may result when there are many middle-income families and relatively few very rich ones. They use a statistic called the "Gini

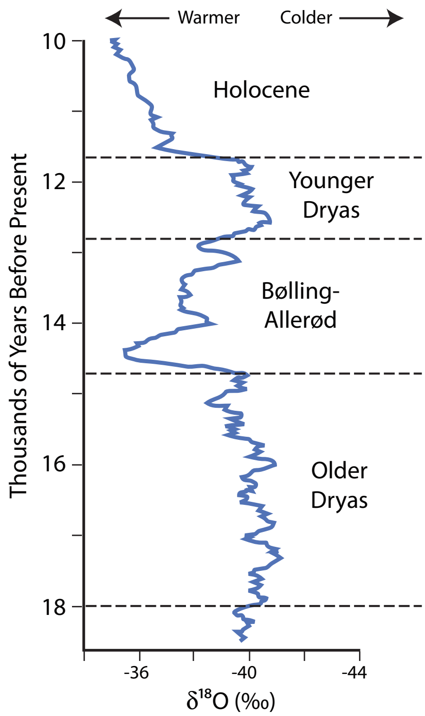


Fig. 5.20 Specialized time-series with bottom-to-top orientation to plot oxygen-isotope ratios ($\delta^{18}O$ ‰) in the GISP-2 ice core from Greenland as a climate-change proxy. (Curve from Buizert et al. 2014: 1178)

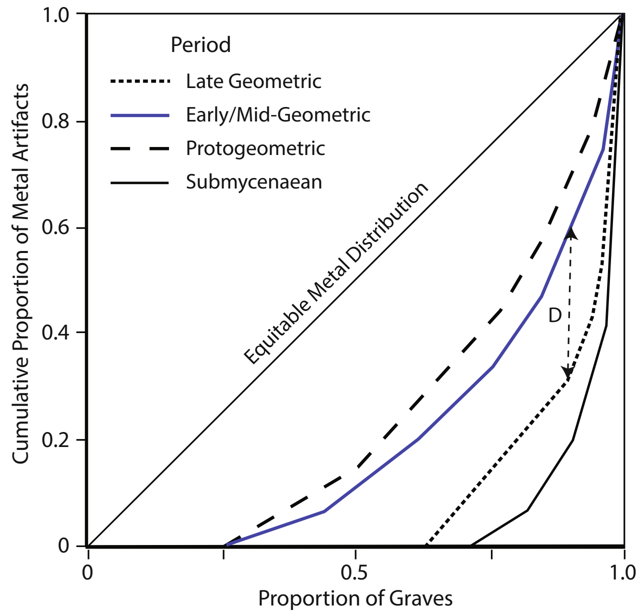


Fig. 5.21 Comparing wealth inequality in graves by means of ogives. The Lorenz curve compares inequality by plotting the cumulative proportion of metal artifacts against the cumulative proportion of graves, ranked from "poorest" to "richest" graves. The diagonal line ($y = x$) would indicate an "equitable" distribution. The curves closest to the lower right corner of the graph show the greatest inequality, with most graves showing no metal artifacts at all (modified from Morris 1987: 142). The maximum vertical difference (D) between curves for early-mid and late Geometric periods provides the Kolmogorov-Smirnov statistic for the difference between these periods' wealth distributions

coefficient" to measure these shapes, and some archaeologists have followed their lead to measure inequality in grave goods (Morris 1987: 141–143).

At one time, the most common archaeological use of cumulative frequency distributions was to compare lithic assemblages (e.g., Bordes 1953: 229). Although plotting two or more of these on one graph does allow easy comparison, this is a misapplication of the ogive, as the x-axis consists of a nominal scale of artifact types in arbitrary order. If you change the order of the types, a completely different shape of broken line results.

5.2.11 Scatterplots

Scatterplots are the most appropriate way to show data along two, or sometimes three, dimensions on interval or ratio scales, and can also be used to compare two or three sets of data on the same graph. They are particularly useful in exploratory analysis (see attribute association, pp. 34), and to find relationships among dimensions.

Most scatterplots have two dimensions, both with continuous interval or ratio scales, that constitute the x- and y-axes of the graph. Each (x, y) point on the graph represents the intersection of the two dimensions in one artifact, or site, or obsidian source, or pottery fabric, for example (Fig. 5.22). The easiest way to add a third dimension to this graph, especially if that dimension is on a nominal scale, is to replace simple points with symbols. For example, a circle could represent a flake and a square represent a core, or a red triangle might represent a chert-tempered sherd while a green triangle represents a calcite-tempered one. One can also add error bars to the points to indicate our degree of uncertainty about the measurement.

If the third dimension is also on an interval or ratio scale, we can represent the three dimensions in computer software as a kind of a box that we can rotate to see how the points are distributed in "space." However, to display this information on two-dimensional media, such as paper, we need to "fold out" the box to show the data from two or three sides (Fig. 5.23). Sometimes, we use thin or dashed lines to join selected points in multiple views, especially they are important, so that it is easier to see how the views are related to one another.

Scatterplots are an excellent way to reveal relationships between two dimensions. For example, if we plot Clovis points' width against length, we might find that the points scatter mostly along a diagonal line; this would indicate the

Fig. 5.22 Scatterplot showing a Factor Analysis (see Chap. 3) of chemical compositions of proto-porcelain sherds from a cemetery and four kiln sites in China. Adding dashed elipses to the graph helps to highlight important clusters in the data. Note the outlier at upper right. (After Yu et al. 2018: 28)

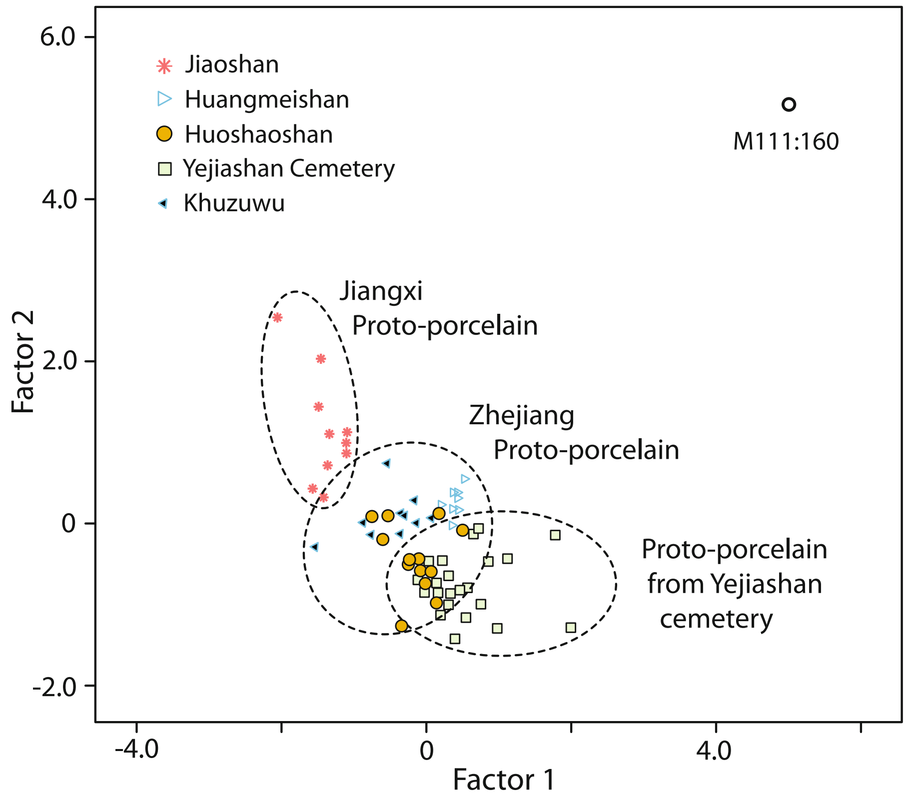
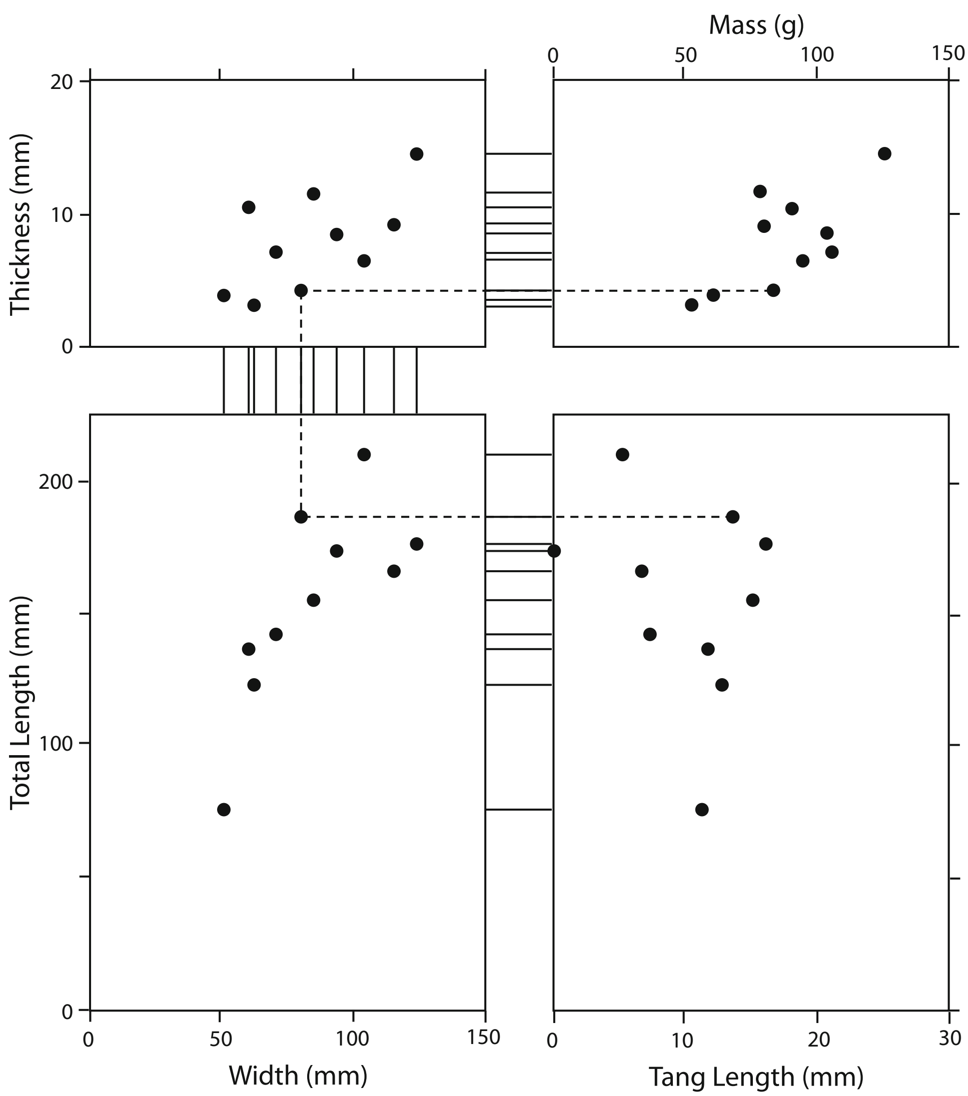


Fig. 5.23 Four-dimensional scatterplot (or "rugplot") with sides of the "boxes" shown side-by-side. The dashed line segments highlight the location of a selected artifact in multiple dimensions, and the short line segments connect the graphs by their shared points (cf Tufte 1983: 135)



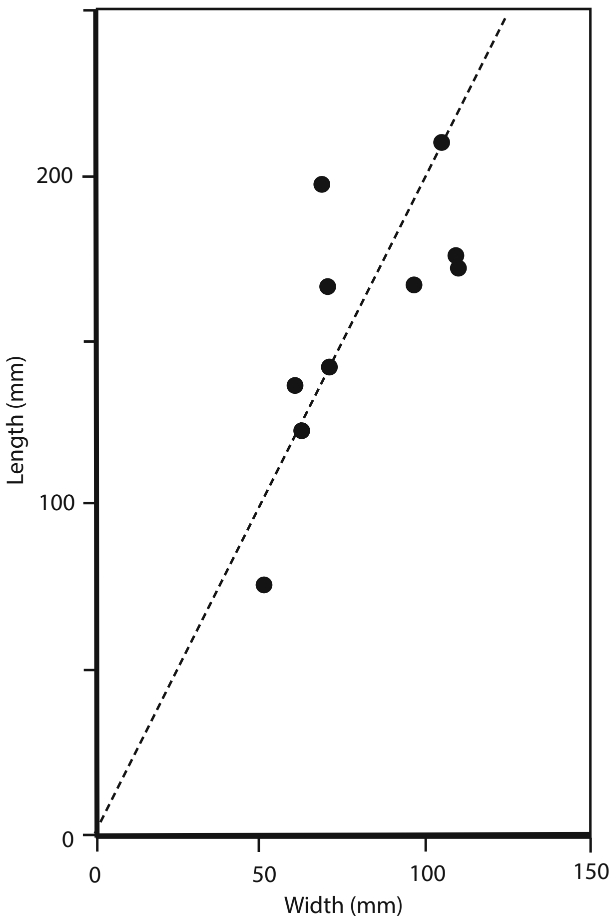


Fig. 5.24 Scatterplot of the lengths and widths of Clovis points from the Hogeye & Graffenried caches in Texas. (Redrawn with data from Waters et al. 2011: 116)

fairly trivial relationship that the flakes are mostly about the same shape, but vary in size (Fig. 5.24), and constitutes a **correlation**. A more interesting example involves an attempt to find a relationship between the number of people who live in a settlement and the total floor area of the settlement as a means for estimating ancient population sizes (Fig. 1.3; Naroll 1962). Archaeologists have interpreted this relationship as a linear pattern with, roughly one inhabitant for every 10 m^2 of floor area (dashed regression line that is roughly $y = 10x$). However, the fact that the points are quite widely scattered and the regression line depends heavily on the largest two or three sites indicates that the correlation is not a very strong one (Brown 1987; LeBlanc 1971; Read 1987: 162).

If the relationship between dimensions might be a causal one, the independent (causal) variable should be on the x-axis and the dependent variable, whose values you hypothesize are affected by the independent one, on the y-axis.

You should keep in mind that a correlation between the variables does not prove causality. In some cases, it is possible that both variables depend on some third variable you did not consider, or the correlation is even coincidental (p. 90).

5.2.12 Ternary Graphs

These are a special kind of scatterplot useful to show data in three categories that add up to 100% (Fig. 5.25). They are commonly used in characterizing soils and sediments (with proportions of clay, silt and sand) but can also be used to show, for example, the proportions of three particularly important animal taxa among faunal remains, or three plant taxa among charcoals or seeds. Points that occur near the apex of the triangle represent assemblages in which one of the three categories dominates, while points near the middle have roughly equal contributions from all three.

Although they look a lot like other scatterplots, they are actually frequency distributions for nominal or ordinal-scale data. They show the same kind of data that you might show in a bar graph or stacked bar graph but are more effective than bar graphs at displaying the relationships among the three proportions when you have multiple cases (e.g., sites or contexts within sites).

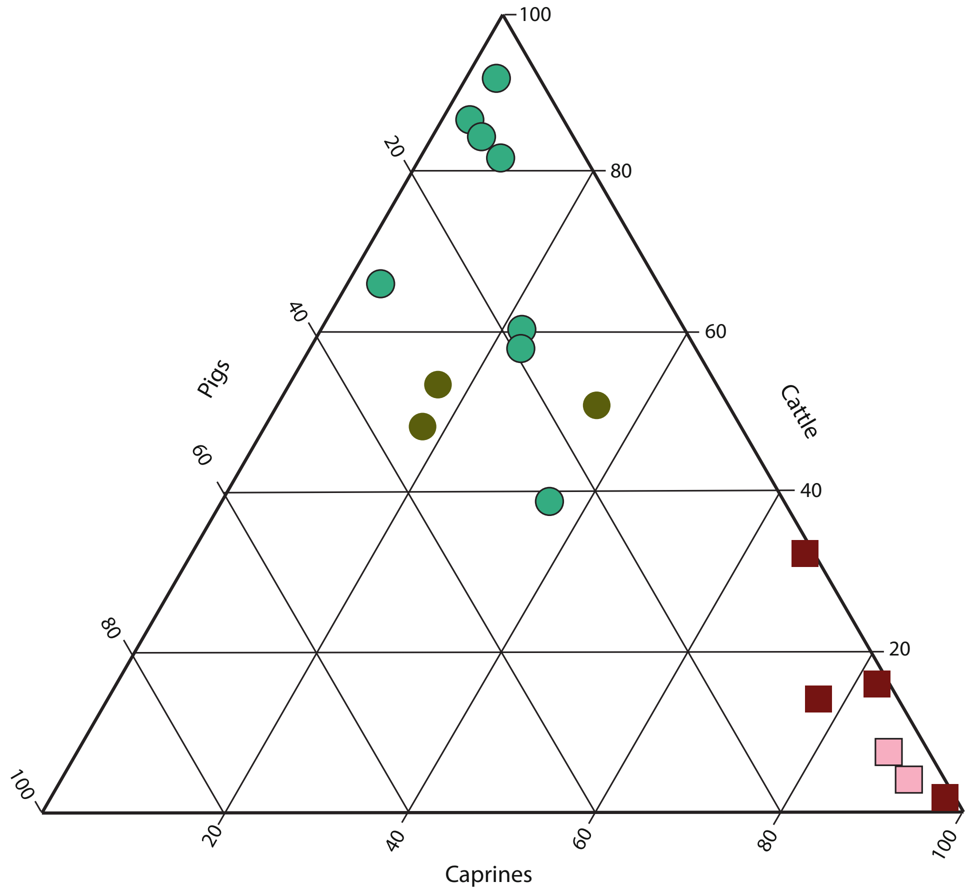
5.2.13 Spatial Histograms, Isopleth and Choropleth Maps

As many kinds of archaeological data have an explicitly spatial component, it is often useful to combine the concept of a frequency distribution with that of a map. In geography, these are called thematic maps.

Choropleth maps are ones in which use of hatching, greyscale or color (or sometimes circles of varying size) serves to differentiate areas on the map that vary in some measurement, such as artifact density (Fig. 5.26a). If, as is often the case, these areas are of unequal size, it is important to use densities, not numbers of artifacts, to avoid distortion. Alternatively, the values could be other direct or indirect measurements, such as a fragmentation index or ratio of burned to unburned bone. As with intervals in histograms, the pattern that appears in this type of map is sensitive to where we put the boundaries between spaces, and how big those spaces are. A key that accompanies the map relates variations in greyscale or color to values on a scale. It is important to ensure that these increments are easy to distinguish, and to think carefully about the choice and order of hues on such a scale so that they intuitively make sense and are easy to distinguish (Monmonier 1991: 150–153).

In spatial histograms (also known as stepped statistical surfaces), the bars of the histogram are three-dimensional, with volumes proportional to the number of artifacts or other countable observations (and heights proportional to artifact density) associated with each space on the map, usually divided into squares of equal area. A square with an unusually high density of pottery, for example, will have a high bar,

Fig. 5.25 Ternary graph of the percentages of domesticated cattle, caprines and pigs at Neolithic sites in the western Balkans (after Orton et al. 2016: 12). Circles represent open-air sites and squares cave sites, and the colors represent different regions. Note that cave sites are high in caprines, and open-air sites high in cattle, while pigs occur more rarely at both kinds of sites



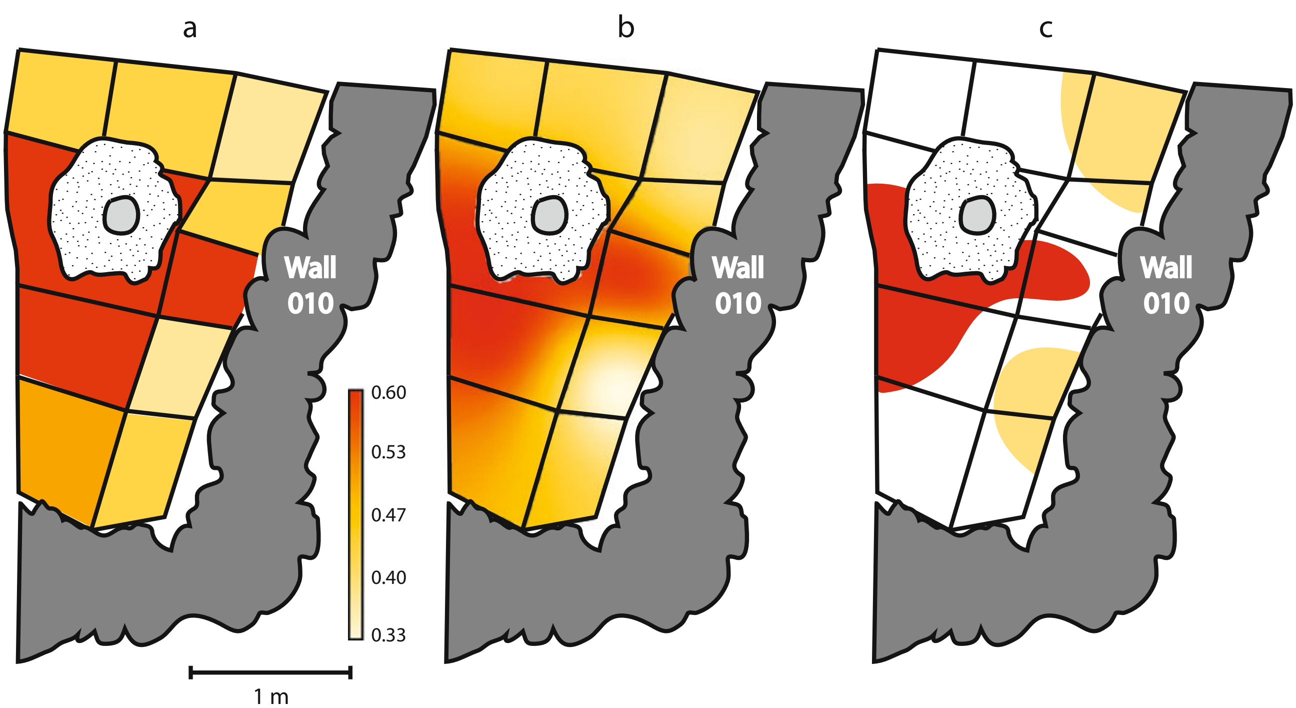


Fig. 5.26 Maps of basalt microrefuse densities on a house floor in a Neolithic site in Jordan (Ullah et al. 2015), with (**a**) a chloropleth map showing number of basalt fragments per cm^2, (**b**) interpolation to smooth the data to make an isopleth map that is not constrained by the grid, and (**c**) with the densities converted to Z scores and areas from (**b**) that are at least 1 Z above the mean in red and ones more than 1 Z below the mean in yellow (see pp. xx). The large object at center is a mortar, distortion to the grid was caused by rocks and other obstructions, and dark grey marks a house wall

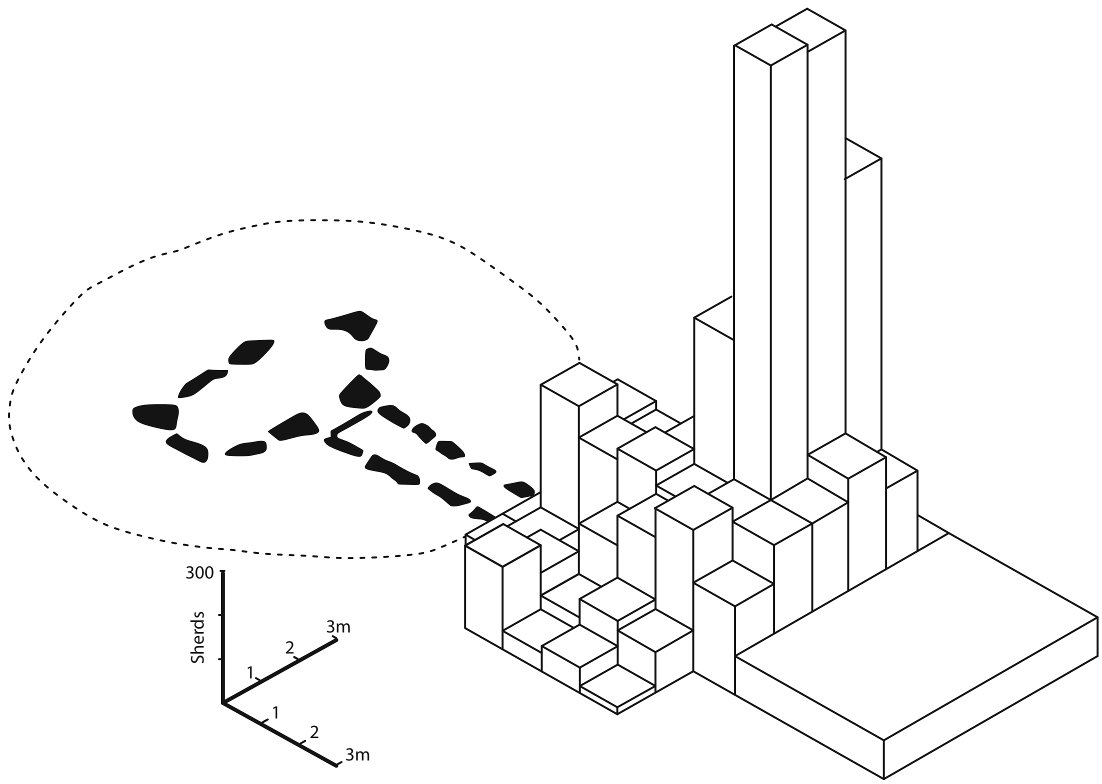


Fig. 5.27 Spatial histogram (or stepped statistical surface) of potsherds at the entrance to Grave 9 at Fjälkinge, southern Sweden (Y. Salama, after Shanks and Tilley 1987: 166). Note that this image exaggerates the density of sherds at lower right. (Because that unit is much larger than the others, it should have been scaled by sherd density, not count)

while one with very low pottery density will have a low one (Fig. 5.27). This allows us to assess at a glance whether or not artifacts or other items tend to be clustered in space. As with regular histograms, the shapes of these plots are sensitive to the size and location of the grid squares that we use as intervals.

An isopleth map takes choropleths and spatial histograms one step further by interpolating between the data points in the grid to create a smoothed surface that, in two dimensions, resembles a topographic map (Fig. 5.26b). In archaeology, the isopleths are curves that usually indicate locations of equal artifact density. When these are close together, they indicate rapid change in artifact density, just as closely contour lines on a topographic map indicate steep slope. Today, there are many software solutions for creating such maps from a grid of discontinuous observations, including spline-tension interpolation and kriging in a GIS (Conolly and Lake 2012: 97–101).

5.2.14 Control Charts

Because ensuring the quality of archaeological data is one of the themes of this book, it seems appropriate to draw attention to a type of graph not commonly seen in archaeology, but that is important in Quality Assurance: the control chart.

Control charts are an important element of statistical process control (Montgomery 2009: 226–343). They are a kind of time-series graph that shows measurements of some quality characteristic as they change over time as well as upper and lower control limits (UCL and LCL), which define the range within which we want the quality to lie (Fig. 5.28). They are more common in industrial applications, where the control limits define the tolerance within which some measurement on a manufactured object must lie. In archaeology, the quality characteristic could be the mean difference or TEM (p. 29) between lab assistants' measurements of, say pottery rim diameter, and the same measurements made by an expert lab supervisor, and the UCL and LCL define the "acceptable" limits within which this difference should lie. In this example, the horizontal center line would be at zero difference because, ideally, there should be no differences between the supervisor and the lab assistants in the measurement of this attribute.

Control charts are useful for monitoring lab processes to ensure that laboratory measurements are within acceptable ranges of error. Because they are time-series graphs, they can

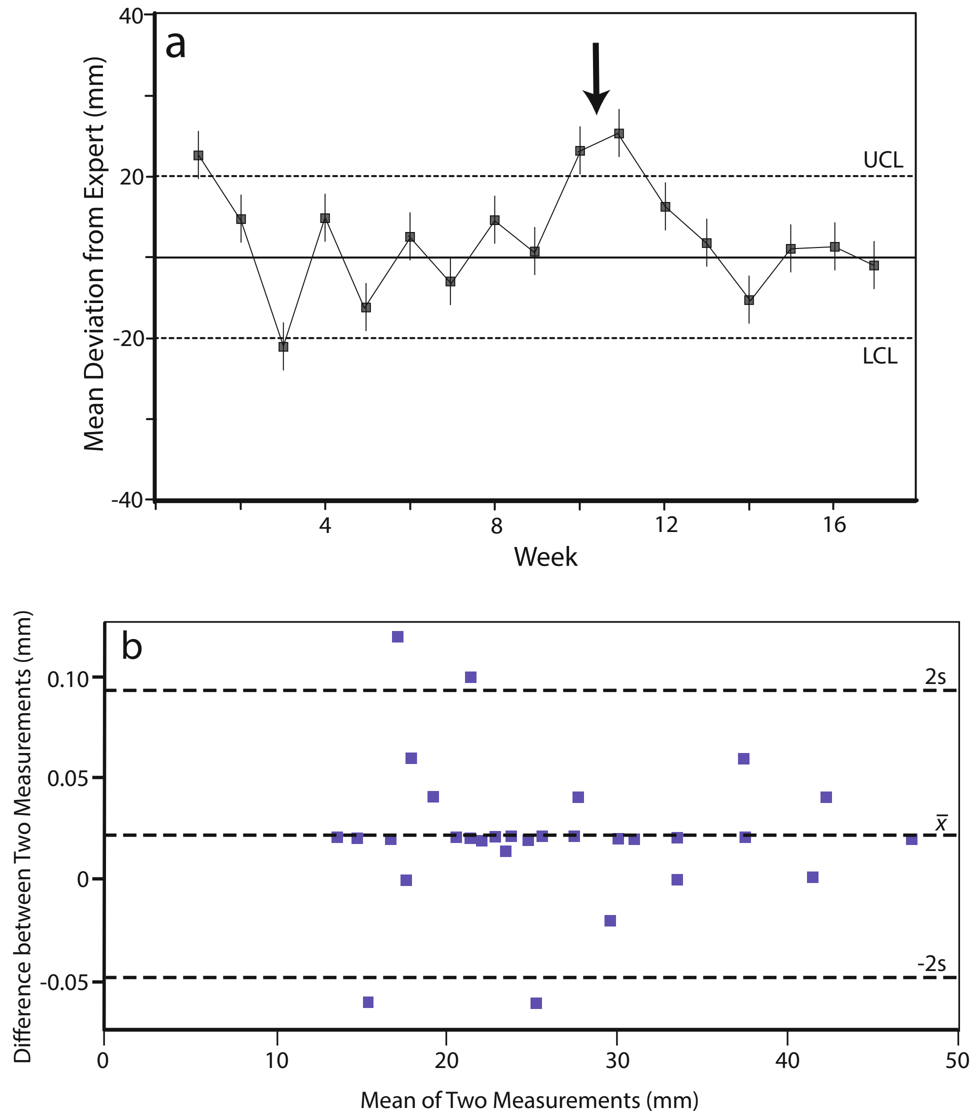


Fig. 5.28 Example of a simple control chart (a), showing Upper (UCL) and Lower Control Limits (LCL) and target value of 0 mean difference between rim diameters by expert and lab personnel. The large arrow highlights 2 weeks where personnel suddenly began to overestimate rim diameter, suggesting perhaps some change in personnel or the expert. Example of a difference-against-mean chart (b) for measurements of projectile point length by the same analyst at two different times (after Lyman and VanPool 2009: 497). As the scatter is horizontal, there is no magnitude-dependent trend, the mean difference is close to zero, and most of the points are within two standard deviations of this mean difference. However, the points are somewhat more widely scattered about the mean for smaller measurements than for large ones

reveal trends or sudden changes in quality. If, for example, the points on the graph trend toward the center line, this would suggest that the lab personnel are becoming more experienced and more adept at making the measurement. If, on the other hand, there is a sudden spike in the graph, this might be associated with a change in lab personnel or lab procedures that call for additional training or review of the procedures to make the lab work more reliable.

5.3 General Principles for Effective Graphs

The bewildering array of possible graph types and software features for "improving" them make it all too easy to create graphs that are confusing, misleading, distracting, or downright dishonest. It is also important to consider your target audience (Wallgren et al. 1996: 64). Will viewers be very familiar with this type of data and its presentation? Or will you need to "walk them through" the data to help them understand it?

Excellent guides for making your graphs more effective and accurate are the books of Edward Tufte (1983, 1990). Tufte emphasizes that graphs are to communicate information clearly, precisely and efficiently. They should make large bodies of data accessible and coherent in a small space, balance fine detail with overall pattern, and encourage people to draw comparisons or identify patterns. Graphs should have a clear purpose and be integrated with their associated statistical and textual presentations. They certainly should not distort data or mislead viewers, but they should also encourage viewers to think about the data's substance rather than the graph itself (Tufte 1983: 13).

Tufte (1983: 53–77, 91–106) identifies graphical integrity and what he calls the "data-ink ratio" as two important aspects of graphical effectiveness.

When he criticizes graphs for having a low **data-ink ratio**, what he means is that the graph is far too cluttered with lines, labels or decorations that are unnecessary for communicating information and can sometimes even distract from that information. For example, putting more increments than necessary on the x- and y-axes of a scatterplot, showing two bars to represent two percentages that add up to 100, redundantly placing values on top of the bars of a bar graph, making the bars on graphs appear three-dimensional when they are not representing three-dimensional data, or decorating graphs with sometimes jarring combinations of color, hatching, or icons, adds to the complexity of the graph without contributing any substance, and sometimes causes distracting moiré effects. Extreme cases of this kind, Tufte (1983: 107–121) calls "chartjunk." Unfortunately, commercially available software makes if all too easy to generate chartjunk. At best, all this does is distract viewers from the information that you really want to convey (see also Lyman & Faith 2018). Keep it simple.

On the other hand, it is also a mistake to remove so much labelling that the viewer has to search through a lengthy caption just to interpret the graph. Include enough labels that it is clear what each axis is measuring, its scale and units, and whether the graph is showing a single set of data or comparing two or more sets. Provide a title, and show relevant information, such as a key, scaling icon, error bars, or one or two descriptive statistics (such as *n*, or sample size), if those are important to accurate interpretation.

Graphical integrity concerns the graph's honesty. Consciously or not, the decisions you make while designing a graph can often make it misleading. One of the most common threats to graphical integrity is to distort its linear scale. This can happen just by hiding the baseline of a bar graph, histogram or time-series of ratio-scale information so that there is no zero at the bottom of the y-axis or a big chunk of the y-axis is missing (Fig. 5.29). This tends to inflate apparent differences between observations. Even worse, some graph authors shift the baseline to a large negative number to smooth out and disguise differences. While there are some instances where it is useful to use a nonlinear scale on one or more axes of a graph, such as a logarithmic scale, to protect graphical integrity you need to make sure that use of a nonlinear scale is very obvious. Note too that this problem does not apply to interval scales that have arbitrary zero, such as temperature or calendar years.

Another common threat to graphical integrity is to ignore the proportionality of area. If you have a graph, such as a bar graph, in which the measurements are proportional to bar height, but you add 3-D effects or, worse yet, icons that are sized to fit those heights (Fig. 5.30), you are distorting the proportionality of the "bars." If one is supposed to show a value double that of another bar, for example, this approach could give it an area not two, but perhaps four times that of the smaller one. Computer software for generating graphs often offers features like perspective views with foreshortening that distort the area of bars or other representations of measurements or shift the origin (zero point) out of view. Since the human eye is heavily influenced by area as a measure of "importance," this is misleading. You should overrule these distortions, and only use 3-D effects to show three-dimensional data, as in spatial histograms, where their use actually contributes to the viewer's understanding of the data.

For graphs that employ color, contrasting hues can help to convey information, or be useless distractions, depending on how we assign them. For example, some graphs and maps usefully employ widespread associations of blue with cold and red with hot to show temperature variation (Monmonier 1991: 152). Similar uses of a scale for choropleth maps with contrasting hues at either end can be helpful when values can be positive or negative but, for densities of artifacts, it is usually better to have only one hue with greater intensity for high values and little or no color for low or zero values. Hues to represent things like color or quality in a pie chart of lithic raw materials, for example, should echo the values they represent, such as dark grey or brown for obsidian, lighter brown or grey for varieties of flint. By contrast, graphs with many bright colors that do not represent anything substantive are distracting at best, while use of some colors, notably red and green, can make the graph useless to color-blind viewers (Lyman & Faith 2018).

Tufte suggests two basic principles to preserve graphical integrity. First, it is better for the representation of numerical quantities to be directly proportional to those quantities. Second, the axes and units should have clear labels on the graph itself, rather than in a caption. Third, the graph should signal important "events" in the data that you want your viewer to notice, such as modes, notable changes in slope, or associations between dimensions. In a scatterplot, for example, you might enclose distinct clusters of points within dashed ellipses (Fig. 5.22).

5.3.1 Transformations and Problems of Scale

Despite Tufte's cautions, there are some instances in which a non-linear scale is preferable or even essential. For example, datasets may include some observations whose values are many orders of magnitude greater than others; this is called skewness towards large values. If we use a linear scale to show these data, small values hug the zero line at the bottom of the graph and it is impossible to see how they vary, while a few very large values are near the top and most of the graph is empty (Fig. 5.31a). One solution to this problem has been to put a break in the y-axis to cut out a chunk of the y-variation, but this certainly violates Tufte's principle of proportionality,

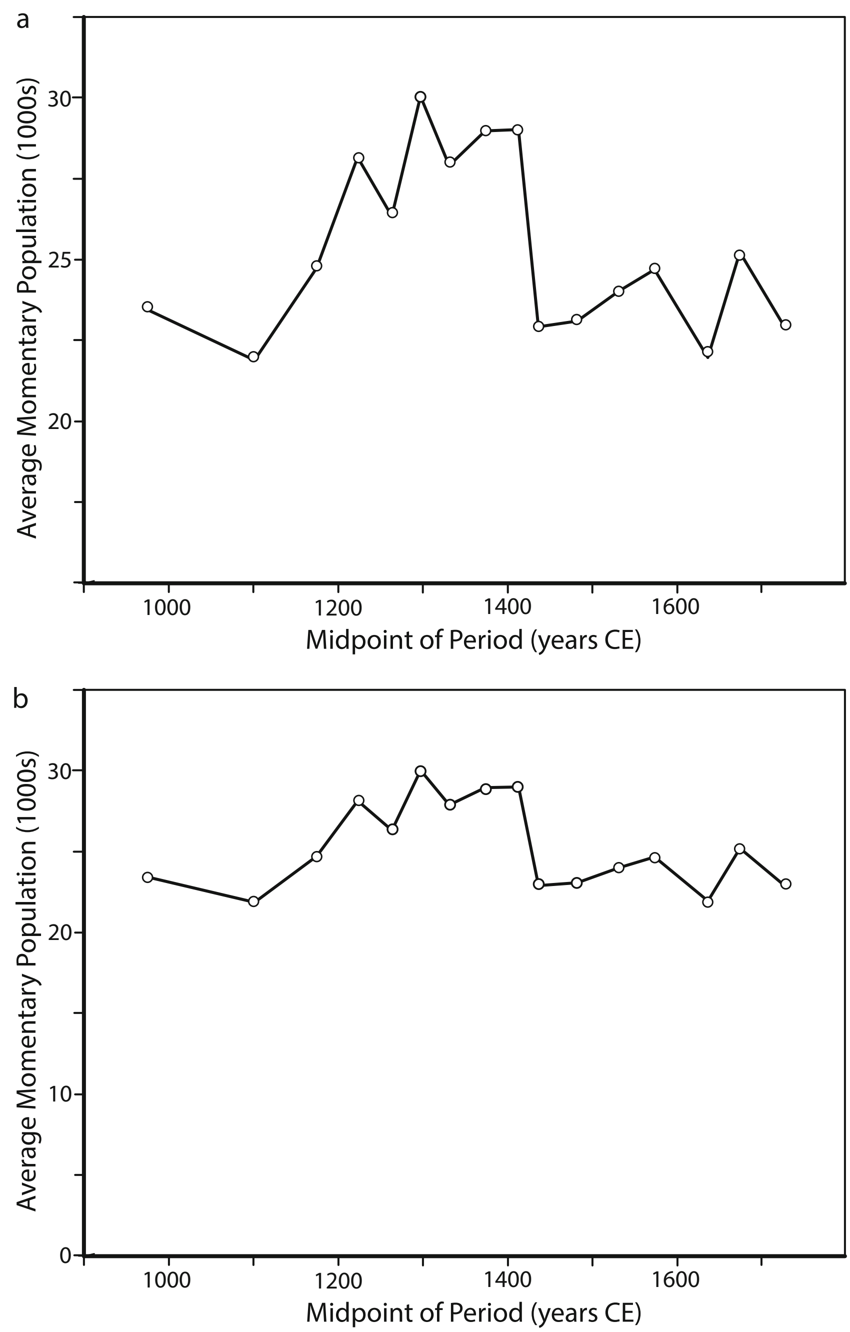
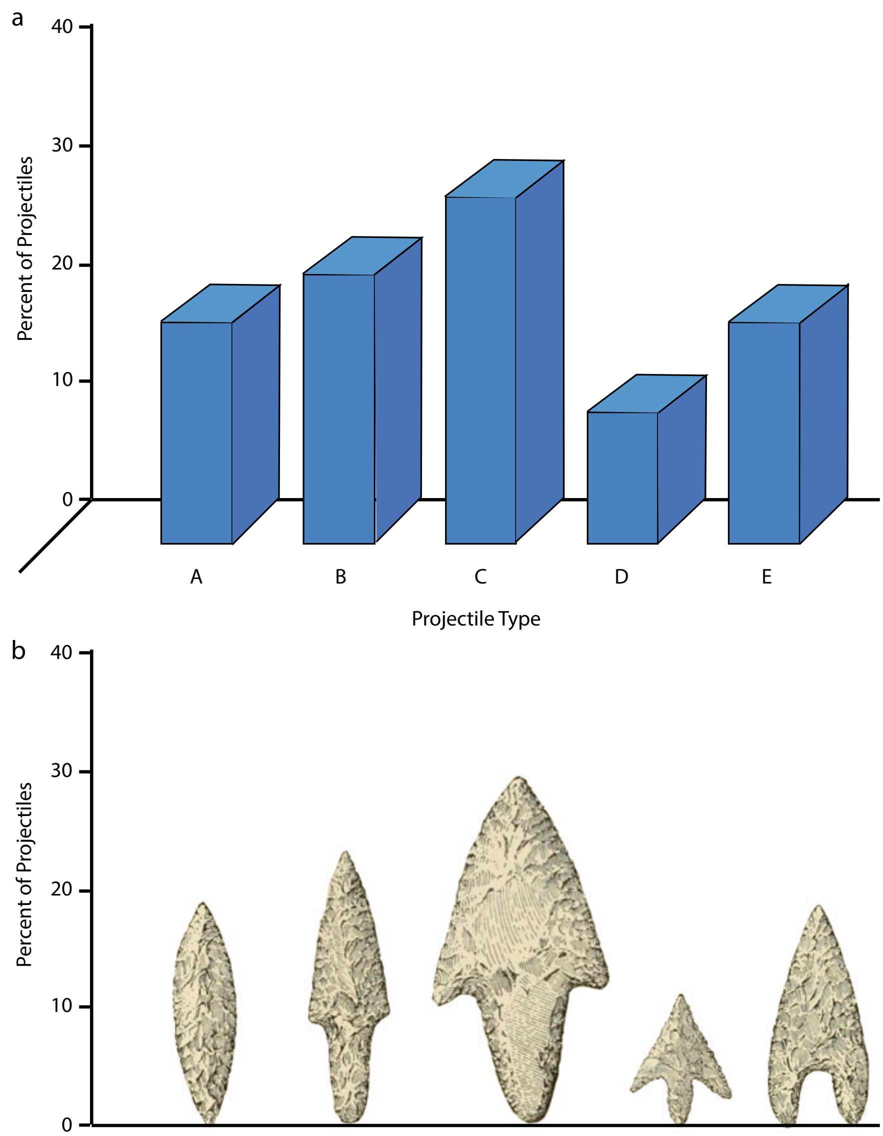


Fig. 5.29 Distortion in a time series (**a**) caused by cutting out the bottom of the y-axis makes the slope of population change appear remarkably steep around 1200 and 1400 CE when an honest graph (**b**) shows that the slope is less pronounced and most of the differences may be within measurement errors

even if you put a jagged mark on the y-axis to indicate what you did.

An alternative is to transform the data. A data **transformation** typically involves using the square root or a logarithm of your measurements, rather than the untransformed data, to make your graphs (Fig. 5.31b). This has the effect of "pulling in" the high values and "stretching up" the low ones. It also sometimes makes patterns in the data more obvious, as in cases where an exponential relationship between dimensions appears as a linear correlation after transformation. In comparing data that have radically different means and dispersion, transforming the data into Z-scores (units of standard deviation) can be useful (Fig. 5.26c). Another kind of transformation that archaeologists often use is ratios in the axes of the graph. For example, they might plot the ratio of two isotopes of Carbon ($^{14}C/^{13}C$) on one axis and the ratio of two isotopes of Nitrogen ($^{15}N/^{14}N$) on the other (Fig. 15.13). Data transformation is also routine in some kinds of statistical analyses, such as Principal Components Analysis (pp. 35). The transformation re-expresses your data in different units. You have to be careful about this. For example, as base-ten logarithmic scales (with increments in powers of 10) have no

Fig. 5.30 Use of meaningless 3D effects (**a**) or icons sized to their heights (**b**) distorts the data by jeopardizing the proportionality of area and also distracts viewers from the data. (Projectile point images from Boule 1923: 378, 380)



zero, you cannot very well use a logarithmic transformation for data that has some zero values. Make it clear to your reader that the data are transformed by labelling the axes and wording the title or caption accordingly.

5.4 Conclusion

Graphics can be extremely effective tools for presenting and helping viewers understand your data. However, careful attention to the graph's purpose and target audience is essential. Are you trying to help viewers make a comparison? Or convince them that there is an important relationship between two variables? Will your viewers want a lot of detail? Or are you only trying to convey the big picture? Whatever the graph's purpose, you should make sure that your viewers' attention is on important features of the data, not on pointless decoration.

As a final note, it is necessary to observe that many of the software solutions for creating graphs mislabel the graph types, offer lots of distracting decoration, and, in some cases, yield completely inaccurate or misleading output. It is better either to choose software wisely (usually statistical packages or dedicated graphing software are better than spreadsheets), or use the software judiciously to create draft versions that you then re-draw in some kind of drawing software, such as CAD (Computer-Aided Design) or illustration packages. In any case, it is best to use something that allows users to edit the graph's fonts, line thicknesses and labels (see Chap. 21).

Table 5.2 summarizes some of the main features of the graphs most commonly seen in archaeological research.

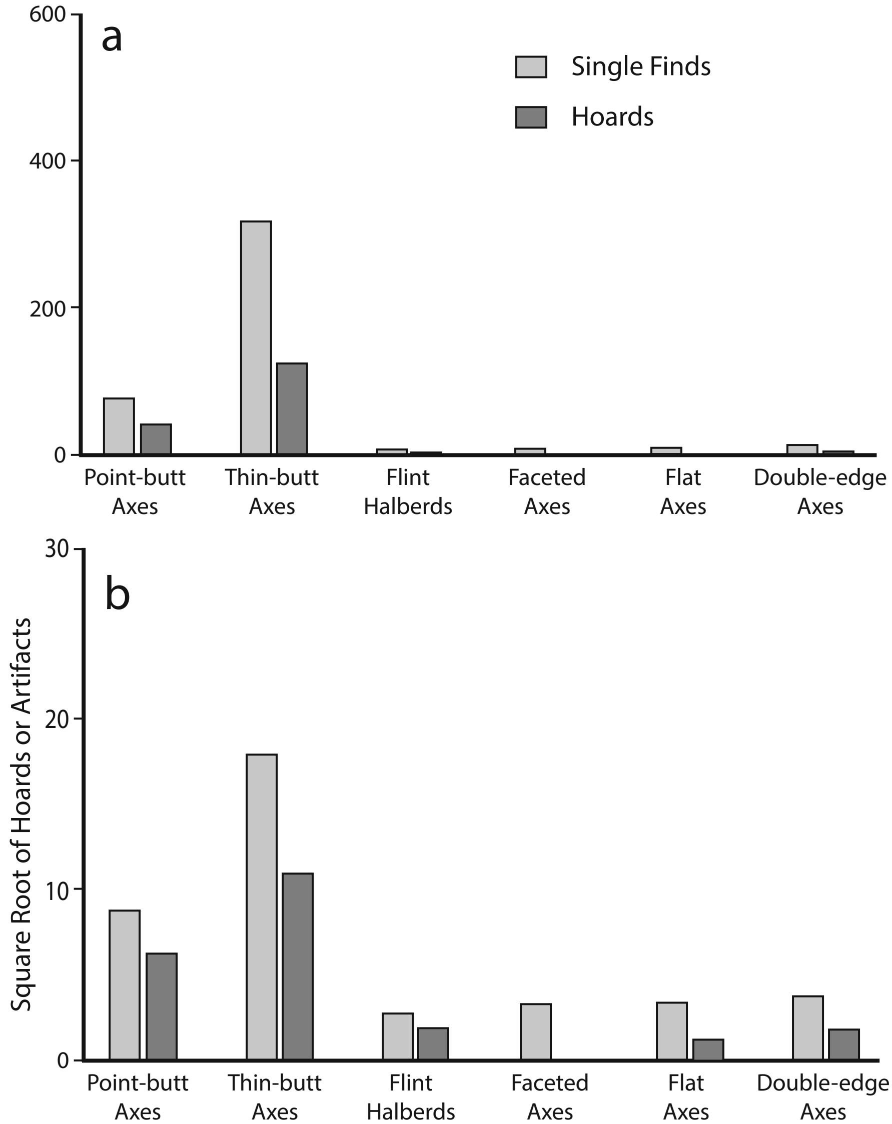


Fig. 5.31 Bar graph of axe types found in single finds and hoards from Middle Neolithic IIB votive deposits at Skane, with a linear scale (**a**) and with a square-root transformation (**b**). The latter violates the principle of proportionality but makes it easier to show both large and small values on the same graph (data from Tilley 1996: 289). Note that a log10 transformation is not possible in this case because there is a 0 value

Table 5.2 Principal varieties of graph and their main features

Type of graph	Number of dimensions	Scales of measurement	Continuous or discrete data	Frequency distribution	Other features
Table	1–3	Various	Both	Can be	Best when small
Box-and-whisker	1	Interval, ratio	Continuous	Yes	
Stem-and-leaf	1	Interval, ratio	Continuous	Yes	Exploratory only
Azimuth plot	1	Interval	Discrete	No	Displays orientations
Bar graph	1	Nominal	Discrete	Yes	
Time-series Bar graph	2	Nominal, interval	Both	No	Shows presence or absence over time
Pie chart	1	Nominal	Discrete	Yes	Useful on maps
Windflower	1	Interval	Continuous	Yes	Displays orientations
Histogram	1	Interval, ratio	Both	Yes	
Line graph	2 or more	Interval, ratio	Continuous	No	Good for time-series
Ogive	1	Ordinal, ratio	Both	Yes	Good for measuring unevenness, comparing samples
Scatterplot	2 or 3	Interval, ratio	Continuous	No	Good for detecting relationships or clusters
Ternary plot	3	Nominal	Discrete	Yes	3 proportions
Choropleth map	1	Nominal, interval	Continuous	Can be	Displays spatial variation
Spatial histogram	2	Interval	Continuous	Yes	Displays spatial variation
Isopleth map	2	Interval	Continuous	Can be	Displays spatial variation

5.5 Summary

- Well-conceived graphs and tables are very effective ways to express complex results simply and to emphasize patterns
- Selection of a table or a type of graph depends on its purpose and the amount, complexity, and scale of the data
- There are many kinds of graphs, but they are not interchangeable, and you should not allow computer software to dictate what kind of graph to use
- Important classes of graph are frequency distributions and time series
- Graphs should highlight the patterns in data or points you are trying to make, rather than distracting viewers with unimportant details or decoration
- Tufte's data:ink ratio is a measure of how well a graph focusses on the data
- However, graphs need to include sufficient information that viewers can discern the graph's purpose and scale or scales, without recourse to a figure caption
- It is very important to ensure the honesty of your graph, particularly by ensuring, whenever possible, the proportionality of area
- Many software packages default to highly inappropriate and sometimes nonsensical graphs

References Cited

Arnold, J. E., & Martin, L. S. (2014). Botanical evidence of paleodietary and environmental change: Drought on the Channel Islands, California. *American Antiquity, 79*(2), 227–248.

Baxter, M. J. (2003). *Statistics in archaeology*. London: Hodder Arnold.

Bordes, F. (1953). Levalloisien et Mousterien. *Bulletin de la Société Préhistorique Française, 50*(4), 226–235.

Boule, M. (1923). *Fossil men: elements of human palaeontology* (J. E. Ritchie, Trans.). Edinburgh: Oliver and Boyd.

Brown, B. M. (1987). Population estimation from floor area: A restudy of 'Naroll's constant. *Cross-Cultural Research, 21*, 1–49.

Buizert, C., Gkinis, V., Severinghaus, J. P., He, F., Lecavalier, B. S., Kindler, P., et al. (2014). Greenland temperature response to climate forcing during the last deglaciation. *Science, 345*, 1177–1180.

Choy, K., Potter, B. A., McKinney, H. J., Reuther, J. D., Wang, S. W., & Wooller, M. J. (2016). Chemical profiling of ancient hearths reveals salmon use in Ice Age Beringia. *Proceedings of the National Academy of Science, 113*(35), 9757–9762.

Conolly, J., & Lake, M. (2012). *Geographical information systems in archaeology*. Cambridge: Cambridge University Press.

Deetz, J. (1967). *Invitation to archaeology*. New York: Doubleday.

De Valera, R. (1959). The court cairns of Ireland. *Proceedings of the Royal Irish Academy, 60*, 9–140.

Drennan, R. D. (2010). *Statistics for archaeologists: A common sense approach*. New York: Plenum Press.

Ehrenberg, A. S. C. (1981). The problem of numeracy. *The American Statistician, 35*, 67–71.

Fladd, S. G., & Barker, C. S. (2019). Miniature in everything but meaning: A contextual analysis of miniature vessels at Homol'ovi I. *American Antiquity, 84*(1), 107–126.

Ford, J. A. (1962). *A quantitative method for deriving cultural chronology* (Technical manual 1, Pan American Union). Washington: Organization of American States.

Gallant, T. W. (1986). 'Background noise' and site definition: A contribution to survey methodology. *Journal of Field Archaeology, 13*(4), 403–418.

Gastwirth, J. L. (1972). The estimation of the Lorenze curve and the Gini index. *The Review of Economics and Statistics, 54*(3), 306–316.

Gelman, A., Pasarica, C., & Rahul, D. (2002). Let's practice what we preach: Turning tables into graphs. *The American Statistician, 56*(2), 121–130.

LeBlanc, S. (1971). An addition to Naroll's suggested floor area and settlement population relationship. *American Antiquity, 36*, 210–211.

Lewis, R. B. (1986). The analysis of contingency tables in archaeology. *Advances in Archaeological Method and Theory, 9*, 277–310.

Lyman, R. L., & Faith, J. T. (2018). Observations on graphing paleozoological data: Suggestions for better graphs. *Geobios, 51*, 435–451.

Lyman, R. L., & VanPool, T. (2009). Metric data in archaeology: A study of intra-analyst and inter-analyst variation. *American Antiquity, 74*(3), 485–504.

Lyman, R. L., Wolverton, S., & O'Brien, M. J. (1998). Seriation, superposition, and interdigitation: A history of Americanist graphic depictions of culture change. *American Antiquity, 63*(2), 239–261.

Makarewicz, C. A. (2013). More than meat: Diversity in caprine harvesting strategies and the emergence of complex production systems during the Late Pre-Pottery Neolithic B. *Levant, 45*(2), 236–261. https://doi.org/10.1179/0075891413Z.00000000027.

McPherron, S. P. (2018). Additional statistical and graphical methods for analyzing site formation processes using artifact orientations. *PLoS One, 13*(1), e0190195.

Monmonier, M. (1991). *How to lie with maps*. Chicago: University of Chicago Press.

Montgomery, D. C. (2009). *Introduction to statistical quality control* (6th ed.). New York: Wiley.

Morris, I. (1987). *Burial and ancient society. The rise of the Greek city state*. Cambridge: Cambridge University Press.

Naroll, R. (1962). Floor area and settlement population. *American Antiquity, 27*, 587–589.

Olson, R. L. (1930). *Chumash prehistory* (Publications in American Archaeology and Ethnology 28). Berkeley: University of California.

Orton, C. (1980). *Mathematics in archaeology*. Cambridge: Cambridge University Press.

Orton, D. (2012). Taphonomy and interpretation: An analytical framework for social zooarchaeology. *International Journal of Osteoarchaeology, 22*, 320–337. https://doi.org/10.1002/oa.1212.

Ortman, S. G. (2016). Uniform probability density analysis and population history in the northern Rio Grande. *Journal of Archaeological Method and Theory, 23*(1), 95–126.

Orton, D., Gaastra, J., & Vander Linden, M. (2016). Between the Danube and the deep blue sea: Zooarchaeological meta-analysis reveals variability in the spread and development of Neolithic farming across the western Balkans. *Open Quaternary, 2*(6), 1–26. https://doi.org/10.5334/oq.28.

Pankowská, A., & Monik, M. (2017). Evidence of innovation and social differentiation in burial practices in Early Bronze Age Moravia. *Archaeological and Anthropological Sciences, 9*(5), 915–933.

Prendergast, F. (2016). Interpreting megalithic tomb orientation and siting within broader cultural contexts. *Journal of Physics: Conference Series, 685*, 012004.

Read, D. (1987). Archaeological theory and statistical methods: Discordance, resolution and new dimensions. In M. S. Aldenderfer (Ed.), *Quantitative research in archaeology: Progress and prospects* (pp. 151–184). Newbury Park: Sage Publications.

Rhodes, S., Banning, E. B., Chazan, M. (2020) Mugharat an-Nachcharini, an Early PPNA sheep-hunting camp in the high Anti-Lebanon Mountains. *PLoS ONE, 15*(1): e0227276.

Sayles, E. B. (1937). Stone: Implements and bowls. In H. S. Gladwin, E. W. Haury, E. B. Sayles, & N. Gladwin (Eds.), *Excavations at Snaketown I. Material culture* (Medallion Papers 25) (pp. 101–120). Globe: Privately printed for Gila Pueblo.

Shanks, M., & Tilley, C. (1987). *Re-constructing archaeology. Theory and practice*. Cambridge: Cambridge University Press.

Shennan, S. (1997). *Quantifying archaeology* (2nd ed.). Iowa City: University of Iowa Press.

Shipton, C. (2018). Biface knapping skill in the East African Acheulean: Progressive trends and random walks. *African Archaeological Review, 35*(1), 107–131.

Singh, P. (1974). *Neolithic cultures of western Asia*. London: Seminar Press.

Tilley, C. (1996). *An ethnography of the Neolithic: Early prehistoric societies in southern Scandinavia*. Cambridge: Cambridge University Press.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire: Graphics Press.

Tufte, E. R. (1990). *Envisioning information*. Cheshire: Graphics Press.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading: Addison-Wesley.

Ullah, I., Duffy, P., & Banning, E. B. (2015). Modernizing spatial micro-refuse analysis: New methods for collecting, analyzing, and interpreting the spatial patterning of micro-refuse from house-floor contexts. *Journal of Archaeological Method and Theory, 22*(4), 1238–1262.

Wallgren, A., Wallgren, B., Persson, R., Jorner, U., & Haaland, J. A. (1996). *Graphing statistics & data*. Thousand Oaks: Sage Publications.

Waters, M. R., Pevny, C. D., & Carlson, D. L. (2011). *Clovis lithic technology: Investigation of a stratified workshop at the Gault site, Texas*. College Station: Texas A & M University Press.

Yu, Y. B., Cui, J. F., & Chen, J. L. (2018). The chemical composition and production area of early Western Zhou proto-porcelain unearthed from Yejiashan Cemetery, Suizhou, China. *Archaeometry, 60*(1), 19–32.

6 Research Design and Sampling

It is the essence of good design that it must be related to the questions asked in individual cases

(Daniels 1972: 201).

At the heart of the research design problem lies bias, and the way in which it enters, or can be prevented from entering, into the data during the research activities of selection, measurement and classification

(Daniels 1978: 29).

Research design provides an explicit plan for ensuring that research will achieve its objectives and lead to valid conclusions (Binford 1964; Daniels 1972; Fisher 1935).

Detailing a research design begins with identifying those objectives, generally through specifying one or more research questions or hypotheses that will focus the research. It also includes reviewing past work that is relevant to these questions, the specific context of the research, and the theories and methods that might be useful in the attempt to answer the questions. These steps help you refine the questions and select an effective approach to the research, sometimes resulting in some very explicit hypotheses that you can test, while at other times just identifying the kinds of data and the methods you will need in order to make inductive inferences relevant to the problem. A research design that uses practice theory as its theoretical orientation should look quite different from one based on phenomenology or evolutionary theory. Further refinement of the research design includes specifying in detail the data needed, how to acquire them, how to ensure the quality of the data, and how to analyze the data to result in an inference or a decision on a hypothesis. Fisher (1935) suggested that it is also necessary to forecast all possible results of an experiment in advance, and to decide "without ambiguity what interpretation shall be placed upon each one of them." This is good advice for at least some kinds of archaeology but is of debatable relevance to others. For archaeologists, the process of research design only occasionally includes experimental design; for most archaeology, we make use of observations on some subset of pre-existing material and, unlike experimental scientists, we cannot repeat our "experiments" or impose controls.

This problem is due to two things. One is that the behaviors and patterns that interest us occurred in the past, sometimes distant past, and we can only observe their traces. We certainly cannot repeat those behaviors under different conditions. The second is that much archaeological fieldwork is destructive. Excavation destroys evidence at the same time that it recovers evidence, so we cannot go back and re-do things if we later think of something we overlooked.

It also brings to mind the distinction between fieldwork and laboratory work in archaeology. While this book focuses on the latter, there is no question that the best research designs include both of them from the start. Jones (2002: 44–50) has noted the serious problem that so often results when post-excavation analysis is distinct from fieldwork. However, this has less to do with an illusion of laboratory objectivity than with a failure to craft a holistic research design in which the laboratory "specialists" contribute to field practices, excavators help set the agenda for lab work, and indeed all the specialists inform and learn from one another. No aspect of archaeological research should occur in a vacuum, and the best archaeology is collaborative and interdisciplinary. However, Jones is also correct that this disconnect often happens. Fieldwork in the heritage industry, for example, sometimes leads to very conventional and standardized laboratory work, and later some other archaeologist might use the resulting data in ways the original

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47992-3_6) contains supplementary material, which is available to authorized users.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_6

archaeologists never anticipated. This almost certainly creates disjunctions that make the second archaeologist's work more difficult, and perhaps less compelling.

6.1 Induction, Deduction, and Abduction

Before getting further into research design, it is worthwhile to review some of the logic of how we make inferences on the basis of data. Terms like "deduction" are so often misused that it is important to understand these clearly. Note, for example, Arthur Conan Doyle's repeated claims that his fictional Sherlock Holmes was making a "simple deduction," when he was actually using inductive or abductive logic.

Valid, non-trivial **deductions** are actually quite rare in archaeology. A deduction proceeds by taking two or more premises that we take to be true, and from them drawing a conclusion that *must* be true, if the premises are true. The most common case of this in archaeology is in stratigraphy (Chap. 19). If we take it as true that layer 7 in a particular site is older than layer 5, and also that layer 5 is older than layer 2, then we can deduce that layer 7 must be older than layer 2. Deduction has a mathematical kind of structure: if $A > B$, and $B > C$, then $A > C$, for example. Most non-stratigraphic archaeological examples of deduction are not terribly interesting: e.g., all Greek amphoras have handles, artifact no. 425 is part of an amphora, therefore artifact no. 425 comes from a vessel with handles. This example is a common type of deduction in which the first statement is a rule, the second is a particular case, and the third is the conclusion. Notably, the conclusion of a deduction must be true, but *only if* the premises are true. In this last example, it is not much of a stretch to imagine that we could be wrong about artifact no. 425 being part of an amphora, especially if it is only a small fragment. In addition, had the second premise been "artifact no. 425 has handles," it would not be valid to deduce that artifact no. 425 was from an amphora, as there is nothing in the premises to limit the occurrence of handles to amphoras (in fact, many kinds of Greek pottery have handles).

Inductions are the mainstay of archaeological inference. Among the key aspects of inductive inferences is that they are ampliative (their conclusions contain information that is not in the premises) and, consequently, their conclusions are never certain (Salmon 1982: 33). Often, the conclusion of an induction makes some reference to probability; if the premises are true, *probably* this conclusion is true (see Chap. 8). For example, our premises might be (1) experimentally cutting bone with a chert flake with an edge angle of 25°, while holding it at an 80° angle and drawing the stroke always in the same direction creates a V-shaped groove with longitudinal striations, and (2) artifact 25 is made of bone and has cut-marks with V-shaped grooves that have longitudinal striations. This might allow us to draw the conclusion, "Artifact 25 has *a high probability* of having been cut by a chert flake with an edge angle of *about 25°*, held at *something close to an 80°* angle, and *probably* drawn in a single direction. This particular kind of inductive inference is called an analogy, a type of reasoning based on the frequency of shared, relevant similarities among the things being compared. Note all the references to frequency, probability and approximation. We cannot be absolutely certain that the tool used on artifact 25 was chert, let alone that it had a particular edge angle and motion of use; we can only say that what we see is consistent with that interpretation. The probability that this conclusion is correct would depend, not only on the evidence mentioned here, but also on how artifact 25 compares with experimental bones cut with different tools, with different materials, edge angles, and motions of use. In other words, the conclusion might be a good hypothesis, but might not be better than competing hypotheses, and indeed several such hypotheses might fit the evidence equally well (**equifinality** is the phenomenon of different processes leading to the same result). Because statistical tests are all about evaluating the probability that a hypothesis is true, they are inherently inductive and, in fact, Peirce, the founder of semiotics, considered induction to be any inference from a random sample to a population (Burch 2018; and see below). A hypothesis that is "well confirmed" by having a high statistical probability can still be false.

Peirce (1883) introduced **abduction** as a type of reasoning that involves taking observations then finding the simplest and most likely or plausible explanation for them. Returning to our amphoras, it could take the form, all Greek amphoras have handles, artifact no. 425 has handles, therefore, artifact no. 425 is an amphora. This conclusion would *not* be a valid deduction because, as already noted, there is nothing in the premises to limit the occurrence of handles to amphoras. Consequently, we should not be very confident of the conclusion, perhaps rewording it as "consequently, artifact no. 425 *might be* an *amphora*." Like induction, abduction leaves some uncertainty or doubt in the conclusion, and could even be characterized as making an educated guess. Peirce saw abduction as the way to formulate a new hypothesis and, in connection with the amphora example, you can see how the hypothesis that artifact no. 425 is from an amphora could lead to detailed comparisons with known amphoras and other handled and handle-less vessels either to strengthen the original conclusion by induction or to lead to a new hypothesis.

However, modern philosophy generally characterizes abduction as "inference to best explanation" (Douven 2017; Lombrozo 2012). Like induction, abduction is ampliative, in that its conclusions contain more information than the premises. However, unlike induction, abductions are partly based on explanatory considerations — you believe that a particular hypothesis is likely to be true because you can think of a plausible explanation that connects the premises with the conclusion — rather than just probabilities or statistical statements as in, for example, analogies (two things share *a large number* of relevant similarities). Abduction also takes into account that reasoning does not occur in a vacuum but, rather, we consciously or subconsciously choose among competing hypotheses, generally favoring the one that makes most sense under the circumstances or that fits with other hypotheses we accept as true. Abduction is common in everyday reasoning.

6.2 The Scientific Cycle

Some archaeologists liken archaeological research to a cycle (Fig. 6.1) whereby we draw inferences inductively from "facts," initially through exploratory research, to create hypotheses and then, through what processual archaeologists called "bridging arguments," determine the (likely) consequences of those hypotheses being true and attempt to test or verify the hypotheses by reference to new data. In other words, we seek out data that could either support an existing hypothesis or demonstrate that it is highly unlikely to be true. This is actually a simplification of the scientific process because the data themselves are always theory-laden, but it is also true that focusing on trying to verify or falsify an existing hypothesis helps us to specify the kinds of data that we need, sometimes leading to entirely new kinds of data. It is also worth pointing out that the data we collect are selected from a large, or even infinite, set of potential data. Conventionally, we describe the selected data as a sample and the larger set of potential data as a population, a distinction to which we will return in discussion of sampling, below (pp. 94–102).

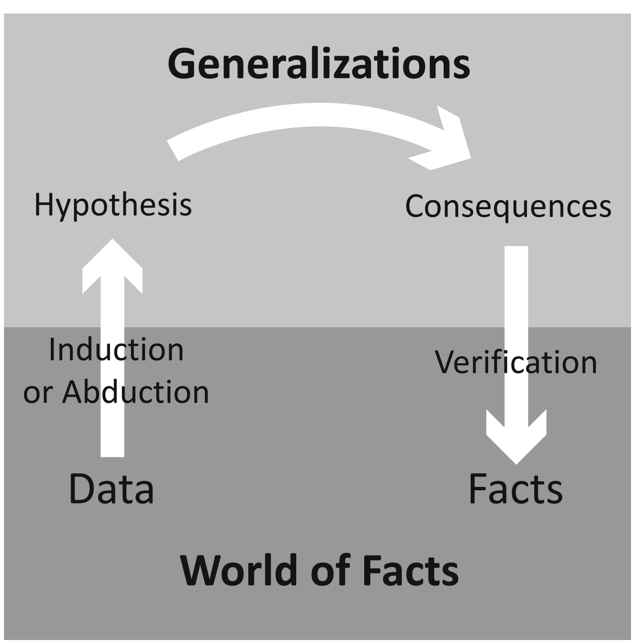


Fig. 6.1 The "Scientific Cycle"

Some people have claimed that the second part of this scientific cycle is deductive because it sometimes involves use of the hypothetico-deductive method of verification. However, it is in fact inductive because the conclusion of verification can be incorrect, even if all the premises are true (see below), and those conclusions, at least in archaeology, are usually probability statements.

Other archaeologists (Johnsen and Olsen 1992) instead make reference to the Hermeneutic circle (Fig. 6.2). This has superficial similarities to the scientific cycle but differs in fundamental ways. This cycle was pioneered in textual studies to enhance understanding of texts whose language was only imperfectly known and that came from an inherently foreign culture, such as the ancient culture of the Kingdoms of Israel and Judah as described in the Bible. Heidegger (1927) and others gave it wider application to gain understanding through a back-and-forth process of attempting to understand the whole from its parts and the parts from the whole. Many archaeological adherents of hermeneutics subscribe to Gadamer's (1975) version whereby the parts-to-whole interplay is one between an interpreter and a text (i.e., the archaeological "record").

Yet another model may describe what archaeologists do, particularly in the face of incommensurable theories or in trying to understand an extinct culture. Bernstein (1983: 69) observes that researchers cannot make "a linear movement from premises to conclusions or from individual 'facts' to generalizations" but must instead follow "multiple strands and diverse types of evidence, data, hunches, and arguments." Alison Wylie builds on this to argue that our theoretical positions do not entirely determine either our interpretation of evidence or formation of hypotheses, and that we regularly are surprised: "we can 'dis-cover' things we did not or could not expect... We frequently find out that we were wrong, that the data resist any interpretation that will

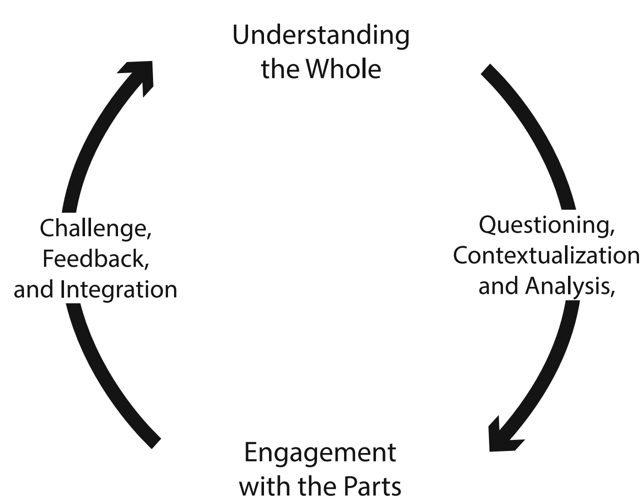


Fig. 6.2 The Hermeneutic Circle

make them consistent with our expectations" (Wylie 1989: 16). This model uses a metaphor of "tacking" back and forth between "cables" that represent different sources of evidence and multiple strands of argumentation.

A good deal of archaeological work is mainly inductive in the sense that archaeologists conduct fieldwork, sometimes in a heritage management context, that may have only very general "problem domains," often including ones that focus on evaluating the significance (i.e., "value" for preservation) of archaeological traces. In some disciplines, this is called a "case study" or a descriptive design. Researchers in this context may still make inferences that lead to entirely new and sometimes surprising hypotheses. However, some other researchers may later revisit these hypotheses and re-evaluate them in a way that resembles the scientific cycle.

Multistage research designs are ones that start with exploratory or "pilot" research that specifies problem domains, refines research questions or, most commonly, provides preliminary information on the character of the data. The next stage takes advantage of these preliminary results and refinements to conceive of a more effective research design.

6.3 Problem Domains

Problem domains, through review of known or suspected information or current debates and available methods, help you select not only what data to collect but also how much data and the methods and strategies to collect and analyze them. You will see examples of this process throughout this book. For example, the nature of your research problem and the context of research will affect your decision on whether to have counts or ubiquity measures (see Chap. 7), whether to take a probability sample or a purposive one, or whether to avoid washing artifacts so that you can do chemical residue analysis or starch analysis. For research designs that include statistical analyses, problem domains also help us specify the "population" of interest, and decide how much error is tolerable, how big a sample to use, and how much confidence we should have in our result (see Chap. 8). Thoroughly documenting problem domains helps you focus on a single research paradigm for your project.

In most places in this book, "population" refers to a target group of sites, artifacts, sediments, or some other thing that we cannot, or should not, examine in its entirety, so that we only examine some subset called a "sample." This statistical kind of population may differ from what biologists sometimes mean when they say, "population," because they are not necessarily thinking of it statistically. Carefully specifying a population in ways that make it archaeologically accessible is an important aspect of problem domains (see Sect. 6.6, p. 108).

Commonly, however, the "population" we are able to study is a collection from fieldwork already conducted, often by someone else. It would have been preferable to have included the fieldwork in the research design, but here we have no choice but to make use of artifacts or other materials that resulted from someone else's research design. In such cases, it is important to review carefully what that design was, what was the purpose of the research, what methods it involved, and other factors to identify possible biases and ensure that it will provide a valid basis for a new research program.

In well-organized archaeological laboratories, there are explicit protocols that outline how artifacts and other materials will be processed, how measurements should be taken and verified, what kinds of analysis will be carried out, and how the data will be stored and maintained (usually on computers). These are informed by the lab's research priorities and other aspects of problem domain, and are also an aspect of Quality Assurance.

6.4 Types of Research Designs

It is possible to classify research designs in a number of ways, and the categories outlined below are not mutually exclusive. For example, exploratory research can also be historical. One of the key features of a research design, no matter to which type it belongs, is that it reasonably ensures the validity of any conclusions based on the research.

6.4.1 Exploratory and Descriptive Research Designs

Some archaeological projects involve compiling data to answer a very broad research question with relatively little idea of what the answer will be. In the simplest cases, it can involve collecting quite a lot of data relevant only to very general research questions, such as, "what was people's diet like at this site?" while in others it can be more focused. For example, archaeologists with the broad goal of determining the sources of obsidian, chert, or metal ore used in the manufacture of the artifacts at a site may use an archaeometric analysis, such as X-ray fluorescence or Instrumental Neutron Activation Analysis for the trace elements in the obsidian or chert, or lead-isotope analysis in silver coins, in the hope that the artifacts will cluster into distinct groups on the basis of their elemental or isotopic compositions and that these groups will correspond with distinct sources. The research design is quite simple, but requires, at a minimum, selecting an appropriately diverse sample of artifacts. If the artifacts show no potentially useful patterns in their compositions, then the researcher might conclude that it is probably not possible to identify distinct sources. If there do seem to be some patterns, then it is possible that the clusters might signal different sources. In that case, it is necessary that the research design includes samples of obsidian, chert or metal ore from a large number of known sources, and preferably several from each source. Once these are also analyzed for the same suite of elements or isotopes, comparing the original clusters to the positions of samples of known source, it is often possible to distinguish sources fairly convincingly. It is an inductive research strategy because we infer that particular clusters may correspond with particular sources, but we can never be certain of those conclusions. There could be other sources, not included in the sample, that have closely similar compositions to one or more sources that are in the sample. For metal artifacts, it is also possible that metalworkers mixed metal from different ore sources in a way that coincidentally mimics another source.

Exploratory or descriptive research is often a precursor to more explicitly focused research designs that take advantage of the preliminary data to form specific hypotheses, determine which variables are most likely to be promising, or determine adequate sample sizes. When that happens, it is common to exclude the data from the pilot research from subsequent analysis in order to ensure its independence. For example, a researcher might use half the data from an archaeological survey to build a predictive model in a GIS, and then test the model using the other half of the data.

However, descriptive research is also often necessary simply to allow another type of research to take place. For example, a historical or causal research design often requires a chronological framework, so it may be necessary to carry out descriptive research on site stratigraphy or radiocarbon chronology even before the other research can start.

One kind of descriptive design, the **case study**, is not necessarily exploratory or preliminary, but involves "digging down" into a particular case so as to understand its workings in much more detail than might be possible in one of the other designs discussed here. Case-study designs are common in ethnography and some kinds of ethnoarchaeology, but also in long-term and extensive excavation programs. Projects that excavate at the same site for many decades gain knowledge of that site in great detail, but we cannot reasonably assume that the results are generalizable to other sites or regions. In fact, researchers sometimes select the subjects of case studies precisely for their distinctiveness. Some researchers call this focused sampling (or purposive selection), which involves selectively studying particular cases that we expect to be particularly illuminating or to provide good tests because they are at the extremes of a population (like the largest site or one of the smallest sites in a settlement hierarchy).

Other designs of this type are quite literally exploratory in that they involve carrying out archaeological search and survey to find archaeological resources. Although surveys certainly can have very focused research designs, some of them may have only very general goals and rather standardized methods. Because these designs involve fieldwork more than laboratory research, they are not discussed further here.

6.4.2 Historical Designs

These designs are of course common in historical disciplines, which not only includes history itself but, broadly speaking, even paleontology, astronomy and epidemiology, not to mention archaeology. The defining feature of historical designs is that they depend on the use of evidence from the past in order to make a case for a hypothesis or to refute it. They are not, as often stated, limited to designs that depend on archival or textual data. Their defining characteristic is that, because their data come from the past, they involve events that the researcher could not have observed directly, let alone manipulate experimentally. Historical designs often focus on identifying and explaining change over time. For example, many historical designs in archaeology have recently focused on the question of whether certain Rapid Climate Change (RCC) events could have influenced ancient cultural changes, such as the collapse of civilizations (e.g., Dalfes et al. 1997). Historical research designs often look for trends in certain variables that are considered likely to be important or, as in the RCC research, looking for coincident changes in two or more variables (e.g., sudden change in population density at the same time as a rapid decline in humidity). Research designs similar to the RCC one are both historical and causal (see below).

6.4.3 Comparative or Cross-Sectional Research Designs

These kinds of designs are appropriate for research that does not focus on change over time, that relies on identifying differences that already exist (rather than resulting from an experimenter's interventions), and compares groups that are pre-existing, such as collections of artifacts from two sites that have already been excavated, rather than groups created by random allocation. This is a very common design in archaeology.

In order to ensure a reasonably high degree of validity, these comparative designs should have a probability component. In other words, the observations used should be selected from all the potential observations through probability sampling. However, it is quite common in archaeology for the comparison to be of two populations that are essentially convenience populations: simply the groups of artifacts or other entities that we happen to have available to us as the result of surveys or excavations already completed. In such cases, we do not have the luxury of going back to resample the site or region and must make do with what we have. To protect validity, we should at least try to determine whether the original data collection for any of the groups was likely to have been biased with respect to the variables of interest.

6.4.4 Regression Analysis and Causal Designs

In archaeology, causal designs are usually also historical designs, except in the case of experimental archaeology. Causal designs attempt to answer the question, "did variable *x* cause change in variable *y*?" by identifying an association or correlation between *x* and *y*. However, simply finding correlations or associations is not enough for a convincing explanation, leading to the truism, "correlation is not causation." Just running statistical software in the attempt to find correlations among a large number of arbitrarily selected variables is highly likely to result in some, but it is also likely that most of these will be **spurious correlations**, and thus this would be a poor research design. Spurious correlations sometimes occur because *x* and *y* are actually correlated with some third variable, *z*, that you didn't think was important, or that just didn't occur to you. However, they can also occur just by chance because so many interval-scale variables either increase or decrease over time, quite independently. For example, there is a correlation between changes in the diameter of the stems of smoking pipes in the United States (Heighten and Deagan 1971) and the price of wheat in Münich (Kellenbenz et al. 1977: 218) over the period from 1688 to 1775 (correlation coefficient, r = 0.85, Fig. 6.3). Surely, there is no common-sense causal relationship between these two variables, and it would have very low validity.

To protect against spurious associations, it is important to make sure that causal research designs are based on explicit theories about relationships or potential relationships among variables. This will at least lead to face validity (p. 13). There should also be explicit and careful consideration of variables that might be *z*, that third variable that caused both *x* and *y*. In addition, causal designs should examine, not just associations, but also the time element: if *x* caused *y*, then *x* must have occurred before *y* happened. Consequently, a causal design of this type should include an appropriate dating method that can evaluate whether this time order is true or likely to be true. For the RCC example, this could require careful analysis of radiocarbon dates with enough precision to distinguish the order of the climatic and cultural events with a high degree of confidence, or use of high-resolution stratigraphy.

6.4.5 Analogical Research Designs

This type of research design can be historical or ethnographic, as in ethnoarchaeology, and is distinct from the role of analogy in some experimental designs, or even its use in modern physics and astrophysics (Dardashti et al. 2016; Thébault 2016). The defining feature of this design is that it involves either collection of data from historical, archival, or early ethnographic texts, or making ethnographic observations in the present, that we expect to be illuminating in some way for our understanding of some process, artifact type, practice, or culture in the past. Analogical designs are

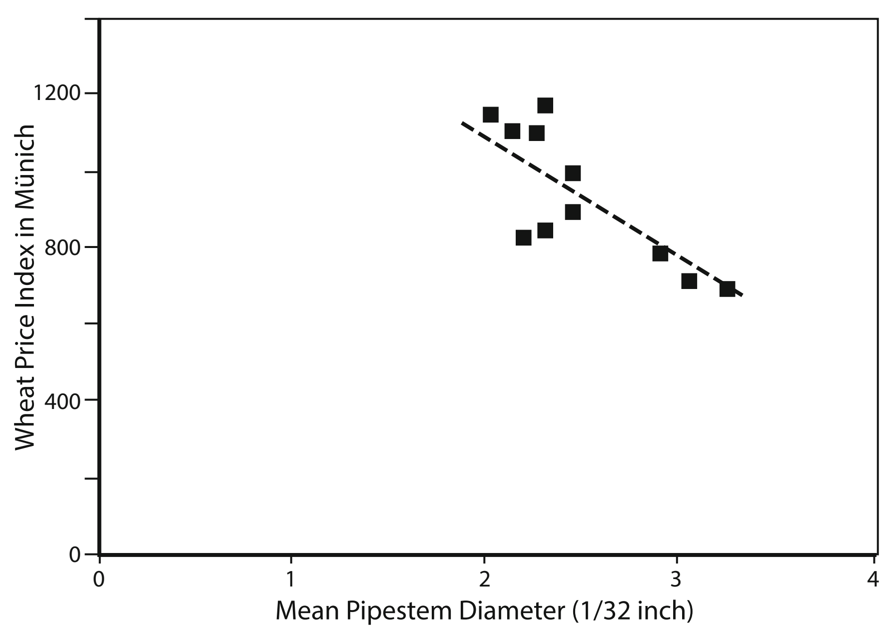


Fig. 6.3 Spurious correlation between changes in the bore diameter of smoking pipes in the United States (Heighten and Deagan 1971) and the price of wheat in Münich (Kellenbenz et al. 1977: 218) over the period from 1688 to 1775 (correlation coefficient, Pearson's $r = 0.85$). The fact that wheat prices happened to rise at the same time that pipestem diameters decreased is purely coincidental

often vulnerable to challenges of their validity; how do we know that the context in the present or in the documentary evidence is sufficiently similar and relevant to the case in the more distant past that we want to understand? Archaeologists who make use of this design have to make this case, sometimes on the basis of a close historical relationship between members of a recent community or cultural group and their likely ancestors (sometimes called the "direct historical approach"), and sometimes by arguing for physical, environmental or practical constraints that would equally have affected both past and recent groups.

6.4.6 Experimental Designs

Experimental research designs are rarer in archaeology than in some disciplines because, like paleontologists, astronomers, and epidemiologists, we typically cannot control or manipulate the conditions that led to the formation of our data because most of those conditions occurred in the past. The exception is experimental archaeology, which often has the goal of simulating some past process in the present, such as flint-knapping, creating wear on tool edges, felling trees with stone axes, or harvesting non-domesticated crops, in order to gain understanding of those processes generally, both past and present. Archaeologists sometimes also employ simulations or re-enactments that they may consider experiments (e.g., Ascher 1961; Ferguson 2010), but they are not experiments in the sense used here (Outram 2008: 2). In "true" experiments, we can impose explicit interventions and experimental controls (Lin et al. 2018). At the same time, much experimental archaeology has an analogical component, in that we make inferences about something we have not observed (because it is in the past) on the basis of correspondence between the relevant attributes of past and present processes that we *can* observe (Salmon 1982: 61–63). Some archaeological experiments involve attempting to reproduce some aspect of an ancient technology and rejecting solutions that do not work; by analogy, we could conclude that they would not have worked in the past either, assuming that the experiment was done correctly. Ones that do work may or may not reproduce the processes used in the past.

Experimental designs permit researchers to have control over at least some of the variables that could help to explain how or why something occurs, in other words, in causal explanation. By manipulating these variables or comparison to "control" groups, researchers can determine the effects of each variable, one at a time or in concert, sometimes with a high degree of confidence.

Experimental designs can themselves be subdivided into many kinds, and selecting which design to use depends on the goals and other problem domains of the research (see below).

6.5 Factors That Can Affect the Validity of a Research Design

While you craft your research design, you should be mindful of the things that could go wrong and try to build in features that protect against them (Campbell and Stanley 1963; Daniels 1972, 1978; Montgomery 2009). These are **confounding** factors, some of them potentially relevant to all research designs, others only to experimental ones. Confounding is an effect due to some unanticipated or uncontrolled variable rather than the variable of interest, and these other variables create competing hypotheses that might explain the effect as well as or better than the hypothesis under study. A good research design should include measures to control for these factors as much as possible or to estimate their probable impact. It is essential to ensure that you can rule out the hypothesis that your results are only due to imperfections in your research methods!

The first seven confounding factors affect the internal validity of experiments and can be factors in both experimental and non-experimental research.

History Conditions can change over the course of the research or an experiment. For example, if there is a change in lab staff or you introduce new training part-way through a project, measurements before and after this change might not be comparable. There could also be gradual change, so slow as to seem imperceptible, in the criteria that lab staff use to classify artifacts or measure attributes (Daniels 1972: 222). In an experiment, it is even possible that changes in lighting conditions or the weather could affect the results.

Maturation Like history, this has to do with the passage of time, specifically gradual changes in the measurers or the subjects in an experiment. For example, subjects doing flintknapping or use-wear experiments get tired over the course of a session, while over longer periods flintknappers often become more skilled.

Testing For experiments that involve multiple tests or questionnaires, the experience of taking the first test can affect the results on the second test.

Instrumentation Changes or differences in equipment or the calibration of measurement instruments, as well as changes in the people doing measurements, scoring or classification, will likely affect results. A common non-historical example in archaeology is the subdivision of an archaeological assemblage into groups that are measured separately by different analysts. Consequently, apparent between-group differences could be due to inter-observer variation rather differences in the assemblage itself.

Regression Toward the Mean This is a problem that occurs when the members of groups to be compared in a second test were selected because of their extreme values on some statistic in an earlier test, creating a tendency toward less extreme values over time. It would not apply to very many archaeological research designs but could be an issue in some quality control studies (see pp. 19).

Internal Selection Bias This occurs when there are differences in the way cases, participants or artifacts are assigned to groups. The most effective way to mitigate this is by random allocation.

Experimental Mortality In studies, especially long-term ones, that involve human subjects, some of the initial participants are likely to drop out over time, or may even literally die, so that the data from later parts of the study are not comparable to those from its inception.

Other factors affect the research's external validity, that is, its representativeness or suitability for generalization. Some of these can affect non-experimental research as well as experiments.

Selection Bias This is a common problem in archaeology whereby a sample under scrutiny is not representative of the population to which we would like to generalize. This can result from non-probabilistic sampling, preservation effects, visibility effects, research priorities and funding, and variations in survey intensity, depth of excavations, or density of remains across a site. Archaeologists have paid particular attention to bias that results from uneven preservation of evidence.

Reactive Effects These occur in an experiment in which training or pre-testing of participants makes them unrepresentative of the population, most of whom did not receive such treatment, but this is an uncommon archaeological research design. More relevant to archaeology are designs in which the set-up of the experiment itself probably makes it unrepresentative of a "natural" setting. For example, in some experiments on archaeological survey, the use of a grid of strings that would never occur in real survey probably had impacts on both survey speed and probability of detection (Banning et al. 2011). Another reactive effect, called the Hawthorne effect or observer effect (McCarney et al. 2007), results simply from the participants' awareness of being test subjects. Again, in an experiment on survey effectiveness, participants tend to walk more slowly and try harder to find artifacts even when they are told to walk at more typical fieldwalking speeds (Banning et al. 2006).

6.5.1 Fisher's Principles of Experimental Design

Fisher (1935) outlined several general principles for the design of experiments that have had long-lasting impact on social and scientific research, only some of which are outlined here.

Randomization In between-group experiments to evaluate the effect of a variable, the subjects should be assigned to each group randomly, while experimental observations that cannot be made simultaneously should be in random order. Randomization distinguishes a "true" experiment from a quasi-experiment or case study. While randomization presents the risk that the resulting groups could accidentally differ in important ways, this risk is manageable with appropriate sampling methods.

Replication The reliability and validity of an experiment can be confirmed when a researcher or research group other than the one associated with the original research repeats it and gets a similar result.

Blocking or Matching Fisher discussed non-random assignments of experimental subjects to groups with the goal of making them similar to one another and reduce irrelevant sources of variation. However, this is at odds with randomization and there are usually better ways to account for possible confounding differences between groups (Campbell and Stanley 1963: 15).

6.5.2 Types of Experimental Design

For more advanced readers, the following briefly outlines some of the experimental designs most relevant to experimental archaeology, but those seriously interested in this topic should consult more specialized literature (e.g., Campbell and Stanley 1963; Cochran and Cox 1957; Creswell 1994; Federer 1955; Montgomery 2009). Most modern experimental designs employ randomization; "systematic designs," by contrast, have disadvantages that have caused most modern researchers to avoid them. These include inability to estimate variance or error, and (spatial or temporal) **autocorrelation** between observations — the tendency for closely spaced observations to be more similar than ones that are far apart (Federer 1955: 11). Much of the literature focuses on the statistics of experiments; here I try to focus more on their structure.

6.5.2.1 Static Group Design

In this design, researchers evaluate the effect of variable X by comparing one group that has been subject to X to another that has not. For example, in a quality evaluation, a lab manager might compare the performance in sorting lithic debitage of lab volunteers who have had a lithics course to ones who have not. This kind of study does not qualify as a "true" experiment because it has no way to assure that the groups are comparable in ways other than X. It is vulnerable to a number of confounding factors, including selection bias (differential recruitment of participants, particularly if one method of recruitment was to make an announcement in the lithics class), maturation (the participants without lithics training might improve their skills at a different rate than the lithics students), and mortality (the lithics students might be more or less likely to drop out than the other participants).

6.5.2.2 Post-test Control-Group Design

This is a simple research design common in experimental archaeology that involves randomly assigning artifacts or research subjects to two groups, giving a "treatment" to only one of the groups, and then measuring some variable to see if there is any relevant difference between the two groups. Random allocation to the groups is assumed to be sufficient for ensuring that the treatment and control groups are comparable (i.e., no selection bias). In experimental archaeology, the experiment might be to evaluate the effect of heat-treating flint, or fletching arrows, or trampling bones.

This design has the advantages of being relatively less costly than other designs, as long as allocation to groups is truly random, is well suited to simple statistical analysis by t-test and, where people are involved, makes it easy to protect participants' anonymity (see research on human subjects, below). It controls for internal validity well but is somewhat vulnerable to mortality effects between allocation and observation. Its external validity, as with other designs, depends on the initial groups being not only equivalent but also representative of some larger population (see Sect. 6.6, below).

6.5.2.3 Pretest-Post-test Control-Group Design

This involves comparing two randomized groups observed before and after a "treatment" that only one group receives, the other being the control group. Thus, it is very similar to the post-test control-group design except that it adds a pre-test, or initial set of observations, in order to be more confident that the two groups really were effectively equivalent before the treatment. In other words, there should be no significant differences between the groups before treatment, but significant differences after it, if the treatment had an effect of the type being tested.

This design controls for most of the confounding variables that affect internal validity reasonably well. Historical events such as having to study for exams (human subjects) or a change in storage environment (non-human ones) should affect both groups equally, as long as they are working or being studied in the same sessions and over the same timeframe. When this is not possible, one can randomize the allocation of group members to time in the lab. Maturation and testing effects should also affect both groups approximately equally. Randomization eliminates the problems of selection bias and regression to the mean, as long as the sample is large enough. It is easy to check whether mortality or the related non-response has affected one group more than the other but, in the event that it has, the two groups will not be strictly comparable.

However, this design is vulnerable to threats to external validity, including interaction and reactive effects. The potential effect of the pre-test arguably makes this design less desirable, at least for human subjects, than the post-test control-group design. And, as with any experiment whose results we would like to generalize, generalization is inductive, so our conclusions could be incorrect even if the experiment was very carefully executed.

6.5.2.4 The Solomon Four-Group Design

This design improves on the simpler control-group designs by explicitly accounting for external, as well as internal validity. It does this by having four groups, two of which are like the two in the pretest-post-test design, and two more that are effectively the post-test design. The former pair has observations before and after a "treatment" but only one of the two receiving the treatment. The other pair is not observed or measured before the treatment, only after it and, once again, one of these groups receives the treatment and the other does not. In other words, the second pair of groups controls for the effect of the pretest. If treatment X really does have an effect, groups that had this treatment should perform differently (e.g., better) than those that did not, not only within their pair, but in the other pair as well. Furthermore, comparison of results from the group that was only measured or observed after treatment X but did not receive it, with the pretest results from the two groups from the first pair allows us to evaluate the effects of history and maturation. However, this design is more costly and complex. Statistical analysis of the results can be accomplished with a 2×2 analysis of variance, with column means estimating the main effect of the treatment, row means the main effect of the pretest, and cell means the main effect of interaction between pretesting and treatment. Where the pretest effect appears negligible, one could use analysis of covariance on the pretest pair, with pretest results as the covariate, to estimate the effect of the treatment.

6.5.2.5 Factorial Designs

These are more complex designs with more groups and more "treatments" or levels of treatment. For example, in a use-wear study, we could use a factorial version of the post-test control-group design, randomly assigning flakes to groups, each of which involved some combination of task (e.g., slicing, carving or scraping), or different tool angles and motions for the same task, and contact material. One group, the control, would not be used at all.

Factorial designs perform better than simple "one-factor-at-a-time" designs because they account for interactions among variables.

Those interested in these more complex designs should consult specialized literature (e.g., Fisher 1935; Kuehl 2000; Montgomery 2009).

6.5.2.6 Latin-Square and Greco-Latin Square Designs

These designs are most common in experiments that have a spatial component, where the goal is to compensate for spatial autocorrelation in the experimental results. Although they can be adapted to avoid temporal autocorrelation, they are probably best used in spatial experiments, similar to those used in agricultural research (e.g., determining the best growing conditions among several alternatives by varying treatments of soils or plants). These designs may have only limited application in experimental archaeology, such as experiments on ancient crop management.

6.5.2.7 Blind and Double-Blind Testing

In experiments like the factorial use-wear one just mentioned, it seems obvious that it would be better if the people participating in the use-wear experiments did not know the details of how they, or the artifacts they were using, were being assigned to groups. This is called a blind test. Withholding this kind of information not only from the experimental participants but also from the analyst is what we call a double-blind test. In double-blind testing, neither the participants nor the people measuring the experimental results are aware of the details of the experiment's structure. For example, in the use-wear experiment, the person examining lithics for traces of use wear will not know to which group each lithic belongs, or even if it was used at all. This prevents participants' or analysts' preconceptions from affecting the analysis of results.

6.5.3 Experiments with Human Subjects

In addition to the concerns discussed above, any experiments that involve human subjects (as opposed to human measurers or observers) have ethical concerns and require approved ethics protocols. Having students help to measure artifacts in your lab is not an experiment on human subjects, but designing an experiment to evaluate the effect of some variable on students' abilities or perceptions is.

A key feature of research ethics is informed and continuing consent, potentially documented in signed release forms.

6.6 Sampling

Protecting validity in many of the types of research design just outlined requires random allocation of cases to groups, or probabilistic sampling of populations. In addition, a large proportion of archaeological research involves statistical analyses that entail the assumption of random sampling. Sampling is less relevant, or even irrelevant, in some kinds of case study in which deep analysis of purposely selected sites or informants provides rich details about a case, rather than making generalizations. There are also many instances of archaeological research in which randomization or probabilistic sampling is effectively impossible (see p. 116). Clearly, these last pose some problems for external validity.

Sampling involves making inferences about a **population** or populations on the basis of subsets of populations. It is important to emphasize that we are talking about populations in a statistical sense, while we sometimes use the word, "population," to refer to groups of people or animals, especially in biological fields, in a way that is not necessarily statistical. We try to estimate **parameters** of a statistical population, such as mean blade length among stone tools or the proportion of deer bones among the faunal remains surviving in a site by measuring **statistics**, such as the sample mean or sample proportion, in a **sample**. To accomplish this with validity, the sample should be representative of the population from which it was drawn. This allows us to estimate the parameters with a particular degree of error and at some selected level of confidence. For example, our research design might call for us to estimate the percentage of deer bones within 5% 19 times out of 20 (or 95% confidence). This statement of error and confidence is critical, for example, for us to tell whether a difference between the sample percentages of deer at two sites is meaningful or just due to sampling error. We also need to be careful to ensure that the population we think we are sampling is actually the one we are sampling. For example, if we are interested in the population of projectile points that were deposited at a site, but collectors have been removing points from that site over the years, the population of points still at the site is not the same as the population originally deposited (Shott 2017).

Probability samples are samples taken from populations randomly or with a method that has some degree of randomization. In **purposive samples**, by contrast, selection of observations is guided by prior information or expert judgment.

To sample a population, we make use of a **sampling frame**, which can be as simple as a list of all the **elements** in the population. In field archaeology, it is often a spatial sampling frame, such as a grid on a map, a set of actual or potential excavation or survey areas, or a list of city blocks in an urban layout. In the lab, it is most likely to be a list of contexts from an excavation, sites from a survey, or bags of artifacts, sediment, or faunal remains from some archaeological field project already completed. Thus, sample elements can be grid squares, sites, excavation areas or deposits, or bags of artifacts. Note that they are rarely individual artifacts, seeds, bone fragments, or the like.

Two other important terms are **sample size**, the actual number of elements in the sample, and **sampling fraction**, which is the proportion of the population included in the sample.

6.6.1 Defining the Population

The population can be finite, such as the population of artifacts sitting on particular museum shelves, or effectively infinite, such as the population of all lithic flakes that a particular knapping technique could theoretically produce. Often, archaeologists would like it to be the population of all artifacts (or lithics, or pottery sherds, or skeletal remains) in a particular layer (or component, or phase, or stratum) of a site but, in reality, we are often limited to the population of artifacts actually excavated from that site. It might be the population of all houses or pits or graves at a site, or only the ones that have been excavated. Commonly it is the population of all (arbitrary) spatial units, such as grid squares, in a region or site. We can even have populations that consist of temporal elements instead of spatial ones, such as the population of levels in a Harris matrix (see Chap. 17). In many of these cases, and especially for populations and elements in space or time, we need to think carefully about their boundaries. Do they correspond with some culturally significant phenomenon, or are they just arbitrary? Accidentally omitting part of a culturally meaningful population can lead to questionable inferences.

6.6.2 Establishing the Sampling Frame and Sample Elements

When the population consists of something like a collection of pots on a museum's shelves, the sample elements just consist of pots and the sampling frame is just a list of all the pots. However, most archaeology is not like that. More often, the population is a spatial or temporal one, and the sample elements are areas of space or moments in time. For example, many archaeologists sample sites or regions by using sampling frames that consist of gridded squares or, less commonly, triangles or hexagons. Spatial elements do not have to be geometrically shaped, however, and could consist, for example, of landforms in a region or rooms in a Southwestern pueblo. Temporal sampling could consist of using periodization, either informal or with the aid of a stratigraphic Harris matrix, to construct a sampling frame that ensured that the sample included observations distributed over a period of time. You should use a sampling frame that makes sense, given your research goals.

6.6.3 Sampling Strategies

When statistical inferences or generalization to some larger population are not needed, you may opt for a purposive sample, which involves consciously selecting certain kinds of observations. For example, some research goals require observations on very rare phenomena that are unlikely to be included in a random sample. Somewhat famously, this would include making sure that research on settlement patterns in central Mexico included the huge site of Teotihuacan (Flannery 1976: 159). However, many kinds of research designs require elements of randomization, and probability sampling of one of the following kinds helps to ensure the validity of research results.

The simplest sample design is the **simple random element sample**. In this design, each element in the population has equal probability of being included in the sample either at every draw (random sampling with replacement) or elements already chosen have no chance to be re-selected (random sampling without replacement, Fig. 6.4a). In the former, selection of any element has no effect on the probability that any other element will be selected, and this is sometimes called **epsem sampling**. "With replacement" means that each element could be selected more than once; "without replacement" means that any element that has already been selected is removed from the "hat" so that it cannot be selected again. Sampling without replacement means that elements selected later on have higher probabilities of selection than ones removed earlier because they are being picked from a gradually shrinking population. Both methods, however, provide reasonably good estimates of population parameters, although not as good as some of the alternatives, as long as we use the correct equations for variance and confidence intervals. For example, Drennan's (2010: 84) *Statistics for Archaeologists* usually assumes sampling with replacement, while Barnett's (1991) *Sample Survey Principles and Methods* assumes sampling without replacement.

Incidentally, if you need random numbers, either for sampling or for experimental design, an excellent resource is the web site www.random.org. This provides true random numbers, rather than just pseudo-random numbers based on a

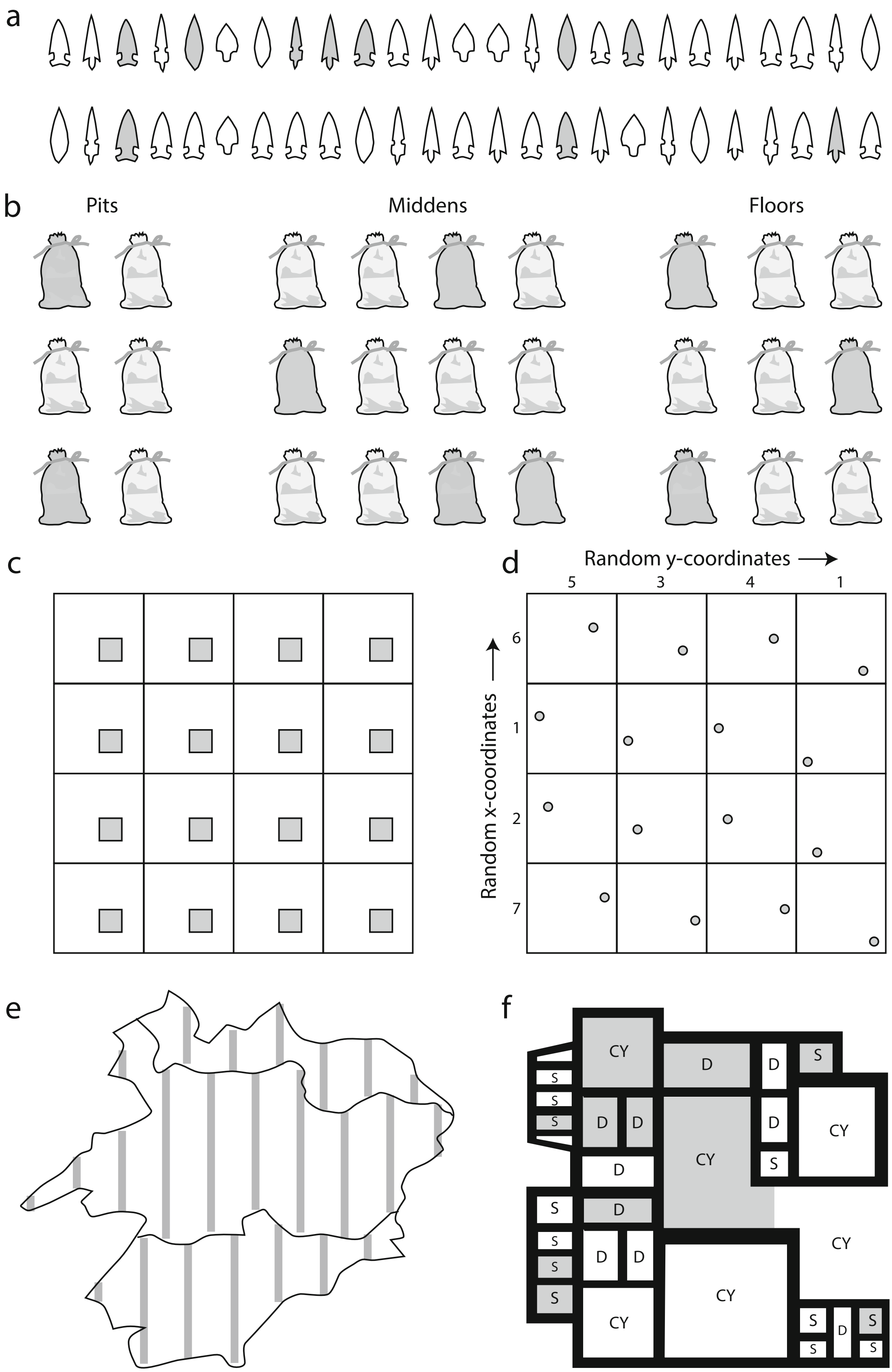


Fig. 6.4 Examples of some basic sampling frames and resulting samples: (**a**) a simple random sample from a collection of projectile points, (**b**) a proportionally stratified cluster sample of artifacts with contexts selected randomly from three kinds of context, (**c**) a systematic spatial sample of squares in a grid, (**d**) a systematic, stratified, unaligned sample of points for coring, (**e**) a systematic stratified spatial sample of transects in an archaeological survey (randomly chosen start transect in each stratum), and (**f**) a disproportionate stratified random sample of spaces in a large architectural complex with three types of spaces (33% CY courtyards, 40% D domestic, and 30% S storage). Selected elements are grey, and sampling was without replacement

mathematical algorithm and is very easy to use. There are also some similar random-number generators that you can find on the web. Otherwise, you can just use a random-number table found in the back of any statistics text, ensuring that you are very arbitrary in your choice of the first number you pick from it and your rule for moving to subsequent numbers.

In **systematic sampling**, only the first element is selected randomly, and all other members of the sample are strictly determined by a regular interval along the sampling frame. For example, we could randomly select one of the first six artifacts in a list of artifacts by rolling a die, then systematically select every sixth artifact in the list until the list is exhausted. Or, in spatial sampling, we could randomly select the first grid square among a set of 1 m × 1 m squares covering a site by using pairs of coordinates taken from a random-number table, and then select every fifth square in a grid centered on that square (Fig. 6.4c). In systematic samples, all elements in the population have equal probability of being selected at the beginning, but once the first is selected, all others are strictly determined.

Stratified sampling is typically superior to random and systematic sampling in that it allows us to account for expected variations in the population and ensure that all those variations are appropriately included in the sample. It involves subdividing the population into subpopulations or **strata**, and then randomly sampling within those. For example, rather than sampling a set of pots as though they are undifferentiated, we might take our list of pots and subdivide it into the categories of jars, bowls, platters, cooking pots, lids, and stands that will serve as strata. Archaeologists have particularly favored stratified sampling for spatial samples in which the strata consist of soil types and the like. However, it is extremely important to ensure that the criteria for stratification are sensible given your research goals and problem domains. For stratified sampling to be effective — and well-designed stratified samples tend to be much more efficient than simple random samples — there should be less variance within strata than between them. In other words, potential observations within each stratum should be relatively similar to one another, but ones in different strata should be rather different, on average. In this way, the stratified sample captures variation in the population quite well.

Stratified sampling can be proportional, meaning that the number of elements in each stratum is proportional to the size of that stratum, or disproportional, meaning that some strata (usually small ones) get a larger sampling fraction than others.

Systematic stratified unaligned sampling is a design specific to spatial sampling that was once popular in archaeology as it ensured fairly even coverage of a site or region without the disadvantage of having such a regular pattern of coverage that would result from systematic sampling. In this design, the space is arbitrarily subdivided into strata by a grid, and then x- and y-coordinates are randomly selected in such a way as to cause all the selected elements in each column of the grid to have the same y-coordinates and each row in the grid to have the same x-coordinates, resulting in a well-distributed scatter with none or almost none of the elements exactly aligned (Fig. 6.4). This does have advantages over simple random and systematic designs but is not as effective as a well-designed stratified sample.

PPS sampling (probabilities proportional to their size) is a specific kind of sampling that involves using randomly located points or line segments to select spaces on a map, inclusions in a thin-section of pottery, or starch grains or phytoliths on a slide (Fig. 6.5). In thin-sections, even the plane of the thin section itself exposes a sample surface that is biased in favor of larger inclusions. Since large spaces, sites or objects are more likely to be selected than small ones

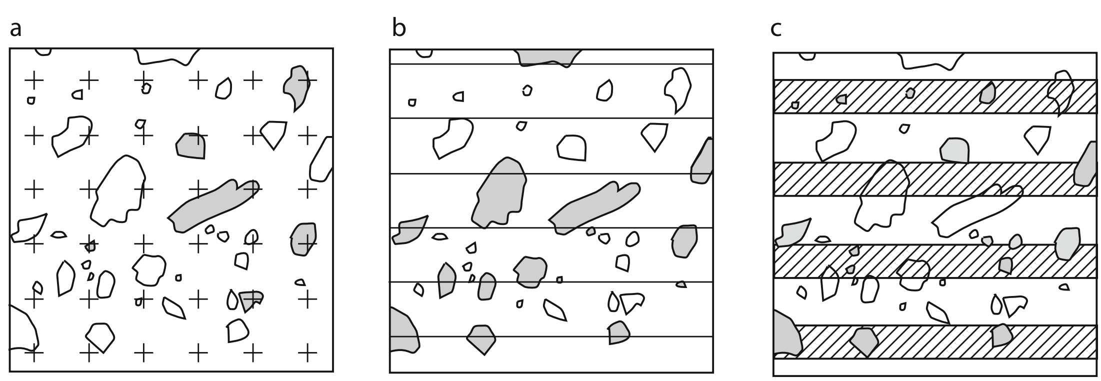


Fig. 6.5 PPS sampling: (**a**) point counting, only the particles or sites intersected by systematic points, (**b**) line counting, only particles or sites intersected by systematic lines, and (**c**) only particles or sites whose lower (or southern) ends are within a ribbon or strip are counted. For (**b**) and narrow particles whose length is less than the line spacing, the probability of intersection is twice their length divided by π times the spacing between lines. An alternative to (**c**) is to count those whose centers are within the ribbon, but centers are not as easy to determine accurately as ends. Note that, for all these methods, large targets are intersected more often than small ones

in this design, it is biased unless we compensate for that effect or are investigating a research problem (such as estimating a ratio between two highly correlated parameters) for which that would be advantageous. Readers interested in this approach should consult more specialized literature (Barnett 1991: 36, 102–103; Orton 2000: 34; Som 1973: 59–74; Thompson 1992: 46–51).

6.6.4 Cluster Sampling

Clearly the easiest way to take a random sample of artifacts or sites is simply to enumerate them and either use a random-number generator to select a subset of them or put the numbers in a hat, mix them around, and pull some out. However, in archaeology, we can almost never enumerate artifacts, features, or sites in advance because we have not found them until after we've done our sampling. Virtually the only times this simple kind of sampling from a list works in archaeology is when we want to sample an existing collection, like the artifacts on a museum's shelves. Instead, we are usually trying to sample from a population of some unknown number of pots, features, buildings or flakes buried in the ground and it is clearly impossible to sample them directly.

Consequently, archaeologists almost never sample pots, or lithics, or even sites. Instead, we usually sample areas or volumes of space, such as a population that consists of 200 m × 200 m squares on the map of a regional survey, or 1 m × 1 m × 0.1 m volumes in an excavation. Less commonly, the sampling element can be temporal instead of spatial, such as a half-hour increment of artifact collection during survey, or 15-minute increment of flake production during a flintknapping experiment. It is incorrect to assume that you have sampled sites, artifacts, bone fragments and charred seeds if you have actually sampled from a population of areas, volumes or time increments.

Using spatial or temporal sample elements in order to sample these smaller things that might be in them is called **cluster sampling**. It is almost certainly the most common type of sampling in archaeology, even if it is not always recognized as such.

In cluster sampling, instead of making a random selection of all members of a population, we concentrate on a spatially or temporally restricted subset of that population that we can reasonably expect to be a "microcosm" of the whole population. For example, we might expect the excavated exposures we make at a site to include almost all the variation we would likely have seen had we excavated the entire site, often on the assumption that the potential observations in each cluster are well "mixed," and thus our statistics should come close to the site's parameters. Archaeologically, the clusters tend to consist of excavation areas or survey transects. To put this in more statistical terms, cluster sampling works well whenever the variance in the clusters (the subpopulation we actually examine) is large relative to the variance in the population. Under those conditions and with roughly equal numbers of observations in each cluster, cluster sampling is more efficient (i.e., has less sampling error) than simple random sampling.

In a sense, cluster sampling involves two kinds of populations. If N is the population of sampling elements (e.g., a set of 1 m × 1 m squares), we also have M, the population of, say, rim sherds that occur in that population of squares. The number of observations in our sample is *n* squares in this example, and these squares contain *m* rim sherds. We are cluster sampling whenever the population and elements of interest (rim sherds, seeds, etc.) are not the same as the population and elements (e.g., site area and excavation squares) that actually constituted our sampling frame.

In typical archaeological cases, the number of observations in each cluster varies. In one excavation square, for example, we might find 50 rim sherds, but only five in others. This can result in biased estimates of population means and proportions, such as mean diameter of pots or proportion of terra sigillata pottery, and their standard errors, if we do not use formulas for these estimates that are designed for cluster samples (Drennan 2010: 243–254; Orton 2000: 32, 212–213; Thompson 1992: 115–118).

Fortunately, the best estimates of means and proportions in cluster samples are the same as for simple random samples. The mean is just the sum of all the measured x_i values in all the clusters divided by the total number of items (m) in all the clusters:

$$\bar{x} = \frac{\sum x_i}{m}$$

This is called the cluster mean. The proportion belonging to a particular class or type i would simply be the total number of that type, summed over all the clusters, divided by the total number of all items:

$$p = \frac{\sum m_i}{m}$$

This is called the cluster proportion.

The approach you should *never* do is to average the individual proportions or densities in all the clusters. Even though each cluster is itself a sample of the population, the mean of all these values is a biased measure of the population mean or proportion.

While calculating means and proportions for cluster samples is straightforward, you need to be more careful about calculating variance, standard deviation or standard error. Unbiased estimates of variance for cluster samples is calculated from the sum of the squared deviations from the cluster mean or cluster proportion of all the clusters, divided by the number of clusters (Drennan 2010: 247–253).

Example

Unfortunately, the common software that most of us use doesn't automatically account for cluster sampling. However, we can do this manually in a spreadsheet. In the rather unrealistic example in Table 6.1, we see cluster samples of identifiable bone fragments (NISP) of *Sus scrofa* (pig) and total NISP (i.e., all identifiable bone fragments) from two sites each with ten excavation units ($n = 10$). For each of these units, we can independently calculate a proportion of *S. scrofa.* For example, for Site A, the proportion for the first sample element is $1/7 = 0.14$. As noted above, doing this for all the sample elements and then averaging the values provides a biased estimate of the population proportion (at 0.28 it is not noticeably different for Site A, but gives 0.44 instead of 0.40 for Site B). The correct proportion estimate comes from totalling all the NISP of *S. scrofa* and dividing that by the total for all kinds of bone fragments ($19/68 = 0.28$, for Site A, $38/94 = 0.40$ for Site B).

Now, to find the correct standard deviation, s, for this proportion, we go back to the proportions for individual elements and find the differences between those and the overall proportion. For example, for Site A, we subtract 0.14 from 0.28, which yields 0.14. In the next column (not shown in Table 6.1), we square these differences (e.g., $0.14^2 = 0.0196$ for the first element in Site A), and then take the sum of these squares (0.301 for all the contexts in Site A). Dividing this value by one less than the number of elements (0.30/9) gives us an unbiased estimate of variance (0.033) and taking the square root of this value gives us the standard deviation, 0.18. Dividing this by the square root of the number of elements ($\sqrt{10} = 3.16$), gives us the standard error, 0.058. Since many of our initial values only had one significant digit, we should round the result off to no more than two significant digits (0.28 ± 0.06 for site A and 0.40 ± 0.06 for Site B), or even only one (0.3 ± 0.1 and 0.4 ± 0.1). Notice that the proportions at the two sites do not seem so very different now. Mathematically, we express this process as:

$$SE_{cl} = \sqrt{\frac{1}{n}\left(\frac{\sum\left(\frac{x_i}{m_i} - p_x\right)^2}{n-1}\right)}$$

However, because the number of potential clusters in the population, N, is small, it is sensible to use the finite population correction, in which case the equation becomes:

Table 6.1 Determining the proportions and standard errors for *Sus scrofa* bones and bone fragments among all the bones and bone fragments (NISP) in samples from two sites, A and B. NISP is the number of identifiable specimens. "Finite" indicates use of the finite population corrector for the standard error (for N = 50 at Site A and N = 100 at Site B) to demonstrate its effect

Site A					Site B				
Sample Elem.	*Sus scrofa* (NISP)	Total NISP	Prop. *S. scrofa*	Diff	Sample Elem.	*Sus scrofa* (NISP)	Total NISP	Prop. *S. scrofa*	Diff
1	1	7	0.14	0.14	11	2	4	0.50	0.10
2	1	5	0.20	0.08	12	4	8	0.50	0.10
3	4	12	0.33	0.05	13	4	6	0.67	0.27
4	2	8	0.25	0.03	14	1	6	0.17	0.23
5	1	3	0.33	0.05	15	6	13	0.46	0.06
6	1	4	0.25	0.03	16	7	9	0.78	0.38
7	1	4	0.25	0.03	17	2	6	0.33	0.07
8	1	5	0.20	0.08	18	3	7	0.43	0.03
9	1	12	0.08	0.20	19	4	16	0.25	0.15
10	6	8	0.75	0.47	20	5	19	0.26	0.14
Total	19	68	**0.28**		Total	38	94	**0.40**	
Mean			0.28[a]		Mean			0.44[a]	
s			0.18		*s*			0.19	
SE			0.058		SE			0.157	
Finite			0.057					0.104	

[a]Note that the mean of the proportions is shown here only to show how it may differ from the cluster proportion (in bold) immediately above it. It is a biased measure and should not be used

(continued)

$$SE_{cl} = \sqrt{\frac{1}{n}\left(\frac{\sum\left(\frac{x_i}{m_i} - p_x\right)^2\left(\frac{m_i n}{m}\right)^2}{n-1}\right)\left(1 - \frac{n}{N}\right)}$$

This involves multiplying the squared values in the numerator of the previous equation by the square of the total count of *S. scrofa* multiplied by the squared ratio of clusters to animal bones, *n/m* (10/68 = 0.147 for Site A, 10/94 = 0.106 for Site B), and multiplying the whole by 1 minus the sampling fraction, before taking the square root. If the sampling fractions were 10/50 or 0.2 at site A and 10/100 or 0.1 at site B, this changes the values of SE somewhat, to 0.057 and 0.10 respectively, giving us even less reason to believe that there is a significant difference in the representation of pig at the two sites, since their standard errors overlap (see Chap. 8).

Calculating variance, *s*, and SE for cluster means works just the same way, by totalling the squared deviations of each element from the cluster mean and making the finite population correction, if warranted.

One of the problems with the prevalence of cluster sampling in archaeology is that we almost never meet the conditions for "good" cluster sampling. Rather than having portions of a site or region that are like microcosms of the whole population, we usually expect much the opposite: that there are activity areas, areas preferred for settlement, and other kinds of spatial and temporal patterning that violate the assumptions of cluster sampling. Most archaeological phenomena provide good examples of spatial autocorrelation. What this means is that artifacts, features, buildings, or sites that are close together in space are more likely to be similar to one another than ones that are far apart. In some cases, this is because artifacts were involved in the same activity, such as cooking or flint-knapping, because sherds or bones found close together could have come from the same pot or animal, or because neighboring sites were exploiting the same agricultural soils or water sources. What this has to do with cluster sampling is that the clusters consist of things that are close together, and thus probably display less variation (less variance) than in the population. Under these circumstances, cluster sampling is less efficient (i.e., has greater sampling error) than simple random sampling. On the other hand, cluster sampling is often our only reasonable choice, given the nature of our evidence.

Some types of research call for **multi-stage cluster samples**. In these cases, rather than examine all the objects in the clusters, we only examine a sample of them. For example, having already selected certain areas within a site for excavation, we later select a subset of deposits in each excavation area for flotation to extract plant remains. This would result in *m* bags of sediment from *n* excavation areas, through a process called sub-sampling. Taking the cluster sampling further, each bag of sediment will contain a certain number of sherds, lithics, seeds, bone fragments or other remains of interest. Multi-stage sampling tends to be useful whenever the number of potential observations is very great and analysis costs are also high, as in archaeobotany, micromorphology or ceramic petrography. Subsampling reduces the analytical effort devoted to each cluster while ensuring that all clusters in the sample receive attention.

6.6.5 Effective Sampling Fractions and Sample Sizes

One of the important decisions a sampler has to make is how many elements of a population to sample.

As already noted, the sampling fraction is just the proportion of elements in the population that we include in the sample, or *n/N*. Despite "cookbook" advice to use a particular sampling fraction, such as 20%, there is no magic recipe for this. Similarly, you may see advice to have sample sizes of at least $n = 30$. However, sample sizes as low as 30 may actually provide very poor estimates of population parameters in many instances. Many archaeologists, having noticed the strong contribution of *n* to the value of SE, have largely ignored sampling fraction and attempted to inflate *n* by using sampling frames with very small elements. However, this is actually a poor strategy (Hole 1980; Ullah et al. 2015: 1251–1253). Using a large *n* of extremely small elements for sampling a site or survey region increases the costs of travel and set-up time a great deal. In both site or regional survey and subsampling sediments for microrefuse, it also tends to depress the number of observations (m_i) in each element, potentially with many "zero" values that may result in biased statistics (see Poisson distribution, Chap. 7). You need to have a reasonable balance.

Decisions about sample size fall into two classes: **fixed-sample designs** that involve deciding, in advance, exactly how many elements to include, and **sequential designs** whereby we keep increasing sample size until we satisfy some criterion. The former is more common in archaeology, if only because resource constraints place an upper limit on how large our samples can be (e.g., Drennan 2010: 126–128; Lee 2012; McManamon 1981).

Informed decisions about fixed sample size must be based on either intelligent guesses or pilot research on what the population parameters of interest and their standard deviations are likely to be, as well as our tolerance of error and intended level of confidence for whatever research decisions we make (such as deciding to reject a null hypothesis, see Chap. 8). The remainder of this section is for more advanced readers who would like to make near-optimal decisions on sample size.

For any given sampling frame, we can use a pilot sample, perhaps from a very small random sample, to come up with very rough estimates of such parameters as μ, σ, and perhaps selected population proportions. We can also decide how much error we consider acceptable for our final estimates of these parameters (our tolerances) and how much confidence we would like to specify for those estimates. As explained in Chap. 8, this confidence has to do with our willingness to accept either a "type I error" (rejecting a true hypothesis) or a "type II error" (failing to reject a false hypothesis). In the context here, the relevant hypothesis is that our sample statistic is accurately estimating the population parameter, and we might be more worried about type II errors (having statistics that are giving us very bad estimates of the parameters).

As for tolerance, we can express that either as a relative error (e.g., an error no greater than 5% of the mean or proportion) or an absolute one (e.g., an error no greater than 5 mm). If we favor relative error (r), and use rough estimates of mean and standard deviation ($\widehat{\mu}$ and $\widehat{\sigma}$), with t as the t-value associated with our preferred confidence level (e.g., $t = 1.96$ for 95% confidence and large samples, see Chap. 8), we can estimate an appropriate sample size as:

$$n = \frac{(\widehat{\sigma}t)^2}{(r\widehat{\mu})^2}$$

Typically, we would estimate the parameters μ and σ with the statistics from a pilot sample, $\bar{x}$ and s.

For an absolute error, we follow the same process except that we replace the denominator with the squared size of the precision (d) we are willing to accept (e.g., ± 4 mm in mean length of blades, or ± 5% in the proportion of deer bones):

$$n = \frac{(\widehat{\sigma}t)^2}{d^2}$$

We can also use this method to find an adequate sample size for estimating proportions. Here, $\widehat{p}$ is an estimated proportion based on a pilot sample:

$$n = \frac{(t\widehat{p}(1-\widehat{p}))^2}{d^2}$$

In both cases, our decisions about how much error to accept or how much confidence we will need depends on how we plan to make use of the results. For example, if we wanted to know the proportion by NISP of *Sus scrofa* bones at hypothetical Site A, we could treat the numbers in Table 6.1 as our pilot sample, and substitute 0.28 for $\widehat{p}$ and, if we wanted to know the proportion within ±0.05 at 95% confidence, 0.05 for d, and 1.96 for t. We would then need a sample size of 63 instead of only 10. For an example of this process, see the case study in Chap. 16 (p. 287).

6.6.6 Sequential Sampling

Another way to approach an optimal sample size involves gradually increasing sample size until the sample meets some predetermined criterion, such as reaching a particular relative error (Fig. 6.6) or a plateau in diversity (see p. 124). At its x_i simplest, it involves a "stopping rule:" we keep adding observations to the sample until we encounter some boundary, and then we base our estimates on the total number of observations we have made up to that point. In archaeology, this has generally involved "sampling to redundancy" (Dunnell 1984; Leonard 1987; Lepofsky et al. 1996;

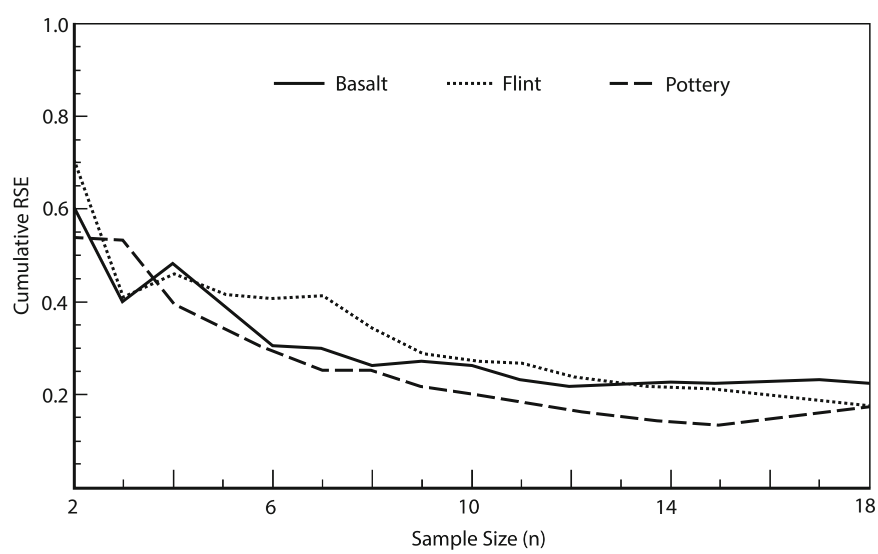


Fig. 6.6 Example of sequential sampling for flint, pottery and basalt microrefuse at the Neolithic site of Tabaqat al-Bûma, Jordan (Y. Salama, after Ullah et al. 2015). After the sample size for these cluster samples exceeds n = 14, there is no significant improvement in the relative standard error

Lepofsky and Lertzman 2005; Lyman and Ames 2004, 2007), increasing sample size until there is no further increase to diversity (often in its simplest version, "richness," the total number of taxa). Truly optimal sequential samples involve more complicated mathematics, and sequential samples are not necessarily less costly than a well-planned, fixed-sample design (Wetherill and Glazebrook 1986: 5, 97–127).

Archaeobotanists and zooarchaeologists have often employed a form of sequential sampling with a stopping rule based on sample richness or a "plateau" in the frequencies or proportions of plant taxa. For example, analysts may keep counting charred seeds or charcoal fragments until they encounter no new taxa in five successive increases to sample size.

6.7 What If Randomization or Probability Sampling Is Not an Option?

It is not uncommon for archaeologists to find themselves needing to analyze data sets, or even collect data, in ways that would violate the assumption of randomization or probability sampling. For example, an archaeological firm might have a contract to mitigate through excavation the impact of constructing an apartment block whose project area overlaps part of a buried village. The contract might call for complete or nearly complete excavation of that part of the site that intersects this project area, but no excavation at all in the rest of the village. There is no reason to be confident that the proposed excavation area will be representative of the whole village; it might be on the village periphery where, say, house-pit density is lower and there are no non-domestic structures. Even more commonly, archaeologists analyze collections of artifacts, faunal or plant remains from projects done years ago and in which they had no part in the research design. Typically, they operate as though the remains available to them constitute some kind of population, but it might not constitute a representative sample of the population that would be of most interest. This kind of situation creates real challenges for the validity of conclusions we might want to draw from these collections or projects, yet there are also ethical imperatives that we try to do so.

6.7.1 Bayesian Sampling and the Concept of Exchangeability

Bayesian approaches (see p. 139) can sometimes provide a way out of this kind of dilemma. Many of the samples that archaeologists find themselves analyzing were not consciously designed as either purposive or probabilistic samples. They may be chance discoveries, samples required by mitigating a development project, "legacy" collections from old excavations, or haphazard samples collected with no explicit research design. Does this mean they are not useful?

A Bayesian would say that they may well be useful. Even if they do not meet the standards of probability sampling, we can assess them for the biases they are likely to entail (Drennan 2010: 88–92). For example, we might expect a group of sherds that someone collected haphazardly from the surface of a Puebloan site in Arizona to have higher proportions of large sherds and painted decoration than we would get from a random sample, just because those sherds are more likely to have attracted the collector's attention. Clearly, any estimates of sherd size or proportion of decoration based on that sample would be biased. However, that does not mean that it is not nearly "as good" as a random sample with respect to things the collector would not have noticed, such as mineral inclusions used as temper or chemical composition of their clay. It is even possible that the collection represents the different kinds of decoration fairly well, although we would want to consider the possibility that the collector favored some designs or was trying to make as diverse a collection as possible.

Rather than worry about the formal properties of the methods used to sample a population, Bayesian statisticians concern themselves with whether there is any *a priori* evidence to suggest that one member of the population is any different from others with respect to the property we will measure or parameter we would like to estimate (Buck et al. 1996: 72–77). If not, it does not matter which members we include in the sample and we can even use the ones that happen to be most convenient, a concept that Bayesians call **exchangeability**. One way to do this is to consider the population to consist of subpopulations, somewhat as in stratified sampling, and the members of each subpopulation as exchangeable. However, in some instances, we would suspect that the criteria for membership in the sample was not independent of the parameter of interest, as in the case of sherd size mentioned above, or collections that came from only one kind of activity area. Because Bayesian statisticians consider all probabilities to be conditional, errors that might be associated with exchangeability may be too small to be troublesome and will be subject to correction in later parts of the research cycle.

6.8 Summary

- Effective archaeological work requires an explicit research design tailored to specific research questions and objectives
- Problem domains include these research questions as well as existing information on the topic, available methods, identification of the target group or population, and

protocols for the measurement, analysis and curation of data

- Research designs can be single-stage or multi-stage, and may conform to one or more of several general types of design, ranging from exploratory and historical to causal, analogical, and experimental designs
- For any research design to be generalizable beyond a case study, it needs to have aspects that protect its external validity
- Confounding factors that can negatively affect both internal and external validity include history, maturation, testing effects, instrumentation, selection bias, and reactive effects. Some of these are specific to experimental designs
- Randomization is the main tool for countering most of these confounding factors and probabilistic sampling is thus an important aspect of archaeological work, especially for avoiding selection bias
- Sampling allows us to make generalizations about a population on the basis of observations on a subset or sample
- Well-conceived samples require informed decisions about sample element, sample size, and whether cluster sampling should be involved
- As so many archaeological samples are cluster samples, it is important to use the correct methods for calculating standard deviations and standard errors
- Where random sampling is impossible, we might use the concept of exchangeability to see if there are aspects of the data that may be similar to a random sample

References Cited

Ascher, R. (1961). Experimental archaeology. *American Anthropologist, 63*(4), 793–816.

Banning, E. B., Hawkins, A. L., & Stewart, S. T. (2006). Detection functions for archaeological survey. *American Antiquity, 71*(4), 723–742.

Banning, E. B., Hawkins, A. L., & Stewart, S. T. (2011). Sweep widths and the detection of artifacts in archaeological survey. *Journal of Archaeological Science, 38*(12), 3447–3458.

Barnett, V. (1991). *Sample survey principles and methods*. New York: Wiley.

Bernstein, R. J. (1983). *Beyond objectivism and relativism: Science, hermeneutics, and praxis*. Philadelphia: University of Pennsylvania Press.

Binford, L. R. (1964). A consideration of archaeological research design. *American Antiquity, 29*, 425–444.

Buck, C. E., Cavanagh, W. G., & Litton, C. D. (1996). *Bayesian approach to interpreting archaeological data*. New York: John Wiley & Sons.

Burch, R. (2018). Charles Sanders Peirce. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.standford.edu/archives/win2018entries/peirce/ (accessed 16 December 2018.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally & Co.

Cochran, W. G., & Cox, G. M. (1957). *Experimental designs* (2nd ed.). New York: Wiley.

Creswell, J. W. (1994). *Research design: Qualitative & quantitative approaches*. London: Sage Publications.

Dalfes, H. N., Kukla, G., & Weiss, H. (Eds.). (1997). *Third millennium BC climate change and Old World collapse*. Berlin: Springer-Verlag.

Daniels, S. G. H. (1972). Research design models. In D. Clarke (Ed.), *Models in archaeology* (pp. 201–229). London: Methuen.

Daniels, S. G. H. (1978). Implications of error: Research design and the structure of archaeology. *World Archaeology, 10*(1), 29–35.

Dardashti, R., Hartmann, S., Thébault, Ki., & Winsberg, E. (2016). *Hawking radiation and analogue experiments: A Bayesian analysis*. http://philsci-archive.pitt.edu/id/eprint/14234. Accessed 23 Nov 2018.

Douven, I. (2017). Abduction. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. https://plato.stanford.edu/entries/abduction/. Accessed 16 Dec 2018.

Drennan, R. D. (2010). *Statistics for archaeologists: A common sense approach*. New York: Plenum Press.

Dunnell, R. C. (1984). The ethics of archaeological significance decisions. In E. L. Green (Ed.), *Ethics and values in archaeology* (pp. 62–74). New York: The Free Press.

Federer, W. T. (1955). *Experimental design: Theory and application*. New York: Macmillan.

Ferguson, J. (Ed.). (2010). *Designing experimental research in archaeology: Examining technology through production and use*. Boulder: University of Colorado Press.

Fisher, Sir R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Flannery, K. V. (1976). The trouble with regional sampling. In K. V. Flannery (Ed.), *The early Mesoamerican village* (pp. 159–160). New York: Academic Press.

Gadamer, H. G. (1975). *Truth and method*. London: Sheed & Ward.

Heidegger, M. (1927). *Sein und Zeit*. Tübingen: Mas Niemeyer Verlag. English edition: Heidegger, M. (1962). *Being and time* (trans: Macquarrie, J. & Robinson, B.). London: SCM Press.

Heighten, R. F., & Deagan, K. A. (1971). A new formula for dating kaolin pipestems. *The Conference on Historic Site Archaeology Papers, 6*, 220–229.

Hole, B. L. (1980). Sampling in archaeology: A critique. *Annual Review of Anthropology, 9*, 217–234.

Johnsen, H., & Olsen, B. (1992). Hermeneutics and archaeology: On the philosophy of contextual archaeology. *American Antiquity, 57*(3), 419–436.

Jones, A. (2002). *Archaeological theory and scientific practice*. Cambridge: Cambridge University Press.

Kellenbenz, H., Schawacht, J., Schneider, J., Peters L. (1977). C. Wilson & G. Parker (Eds.), Sources of European economic history (pp. 190-222). Ithaca: Cornell University Press.

Kuehl, R. O. (2000). *Design of experiments: Statistical principles of research design and analysis* (2nd ed.). Pacific Grove: Duxbury/Thomson Learning.

Lee, G. A. (2012). Taphonomy and sample size estimation in paleoethnobotany. *Journal of Archaeological Science, 39*(3), 648–655.

Leonard, R. D. (1987). Incremental sampling in artifact analysis. *Journal of Field Archaeology, 14*, 498–500.

Lepofsky, D., & Lertzman, K. (2005). More on sampling for richness and diversity in archaeobiological assemblages. *Journal of Ethnobiology, 25*, 175–188.

Lepofsky, D., Kusmer, K., Hayden, B., & Lertzman, K. (1996). Reconstructing prehistoric socioeconomics from paleoethnobotanical and zooachaeological data: An example from the British Columbia Plateau. *Journal of Ethnobiology, 16*, 31–62.

Lin, S. C., Rezek, Z., & Dibble, H. L. (2018). Experimental design and experimental inference in stone artifact archaeology. *Journal of Archaeological Method and Theory, 25*, 663–688.

Lombrozo, T. (2012). Explanation and abductive inference. In K. Holyoak & R. Morrison (Eds.), *Oxford handbook of thinking and reasoning* (pp. 260–276). Oxford: Oxford University Press.

Lyman, L. R., & Ames, K. M. (2004). Samping to redundancy in zooarchaeology: Lessons from the Portland Basin, northwestern Oregon and southwestern Washington. *Journal of Ethnobiology, 24*, 329–346.

Lyman, L. R., & Ames, K. M. (2007). On the use of species-area curves to detect the effects of sample size. *Journal of Archaeological Science, 34*, 1985–1990.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne effect: A randomised, controlled trial. *BMC Medical Research Methodology, 7*, 30. https://doi.org/10.1186/147102288-7-30.

McManamon, F. P. (1981). Probability sampling and archaeological survey in the Northeast: An estimation approach. In D. Snow (Ed.), *Foundations of Northeast archaeology* (pp. 195–227). New York: Academic Press.

Montgomery, D. C. (2009). *Design and analysis of experiments* (7th ed.). New York: John Wiley & Sons.

Orton, C. (2000). *Sampling in archaeology*. Cambridge: Cambridge University Press.

Outram, A. K. (2008). Introduction to experimental archaeology. *World Archaeology, 40*(1), 1–6.

Peirce, C. S. (1883). A theory of probable inference. In C. S. Peirce (Ed.), *Studies in logic by the members of Johns Hopkins University* (pp. 126–181). Boston: Little, Brown, and Company.

Salmon, M. (1982). *Philosophy and archaeology*. New York: Academic Press.

Shott, M. (2017). Estimating the magnitude of private collection of points and its effects on professional survey results: A Michigan case study. *Advances in Archaeological Practice, 5*(2), 125–137.

Som, R. K. (1973). *A manual of sampling techniques*. London: Heinemann.

Thébault, K. P. Y. (2016). What can we learn from analogue experiments? *General Relativity and Quantum Cosmology*. arXiv:1610.05028v1 [gr-qc]

Thompson, S. K. (1992). *Sampling*. New York: John Wiley & Son.

Ullah, I., Duffy, P., & Banning, E. B. (2015). Modernizing spatial micro-refuse analysis: New methods for collecting, analyzing, and interpreting the spatial patterning of micro-refuse from house-floor contexts. *Journal of Archaeological Method and Theory, 22*(4), 1238–1262.

Wetherill, G. B., & Glazebrook, K. D. (1986). *Sequential methods in statistics* (3rd ed.). London: Chapman & Hall.

Wylie, A. (1989). Archaeological cables and tacking: The implications of practice for Bernstein's 'Options beyond objectivism and relativism'. *Philosophy of the Social Sciences, 19*(1), 1–18.

7 Counting Things: Abundance and Other Quantitative Measures

If archaeologists do anything, it is count. We count stones, bones, potsherds, seeds, buildings, settlements, and even particles of earth.

VanPool and Leonard (2011: 1)

Scientists have a laudable but rather Pavlovian urge to quantify, and will often measure and count even when the purpose is not clear.

Hubbard and Clapham (1992: 117)

Once archaeologists have classified archaeological items, a natural step is to count how many belong to each class. Rather than just say that two sites or two layers in a site share some type of pottery or stone tool, or that a site has some unspecified number of deer bones, archaeologists prefer to say how many shared sherds there were or what percentage of the faunal remains consisted of deer. Quantification in this sense is critical to many kinds of archaeological analysis, such as seriating pottery (Chap. 18), studying the "fall-off" of obsidian away from its source, or estimating the contribution of venison to a prehistoric community's diet. But there are other measures that characterize potentially countable things in different ways, such as diversity and ubiquity (pp. 122, 124), and one way to characterize things that really *can't* be counted completely in a meaningful way: fractal dimension (pp. 121).

Typical archaeological remains pose many challenges to meaningful quantification (Baxter 2003: 210–221; Gautier 1984; Orton and Tyers 1992; Lyman 2008; Ringrose 1993a). The fact that they're so often fragmentary is the most obvious problem. While it is easy to count complete pots, blades, bones, and seeds accurately, we have to question the meaning of such statements as "12% of the pottery at the Red Butte site is Kayenta White Ware" because no one counted pots, but only their fragments.

But fragmentation is only one problem. Orton et al. (1993: 209) note that a high percentage of coarse cooking sherds over fine table ware, for example, does not prove that cooking pots were more common than fine table ware in the "life population" of pots in use at the site. Finer pottery could have been treated with more care ("curated"), while cooking pots may have been cheap and expendable or even disposable, so that their breakage and discard rates were much higher. At a given moment in its use, a kitchen might have had ten times as many fine dishes as cooking pots, but the middens could still end up with three times as many sherds from cooking pots as from fine-ware dishes.

Archaeologists cannot completely escape the problem that so many of the things we find are not only fragmentary, they have different rates of fragmentation, different use-lives, and were subject to different disposal practices and different survival conditions.

Since archaeologists deal mainly, not with pots, animals, or even bones of animals, but with sherds and bone fragments, what should they do? Does a sherd have meaning in terms of whole pots? Do bone fragments tell us anything about the relative abundance of animals? Or of meat? Two bone fragments could be from exactly the same bone, or from two different bones in the same animal, or from two completely different animals that may have died at different times. Even stone tools are often broken and pose similar problems (Shott 2000).

The first factor to consider is that the number of countable elements from each whole pot, flake, or animal varies. Some species of animal have more bones in their bodies than others do. Some kinds of pot tend to break into more pieces, on

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47992-3_7) contains supplementary material, which is available to authorized users.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_7

average, than do others. Long blades may break more easily than flakes. Some kinds of plants produce thousands more seeds than others do, while some kinds of seeds or bones represent many times more food than others. In keeping with this book's general theme of measurement and measurement error, it is clear that we need to be very clear about what we really *want* to measure (e.g., amount of food or number of animals or pots?), and its relationship to the indirect measures (e.g., number of flakes, bone fragments or sherds) that we are able to make but that are not really the things of most interest.

Fragmentation is really just one aspect of a more general problem: differential preservation. The probability that anything is preserved after centuries or millennia depends on both its own characteristics and the character of its environment (Chap. 9). Some bones, for example, are highly unlikely to survive trampling or prolonged burial in the ground while others in the same depositional environment survive fairly well. Archaeologists have long known that the density of the bone is a fairly good predictor of this aspect of preservation (Lyman 1984), at least when we control for environmental factors, such as the acidity and salinity of the surrounding sediments. Similarly, the preservation of plant remains depends on such things as whether or not they are likely to be charred, waterlogged, or in a very dry environment, and whether humans were interested in plant parts, such as seeds, that have some likelihood of survival, or leafy parts that they usually consumed completely or that would not survive burial in most environments. Study of these kinds of issue is called taphonomy (see pp. 242–243).

Taphonomic difficulties such as these have led some archaeologists to settle for what they sometimes call "presence analysis," simply noting whether certain classes of material occur or not, and either limiting quantification to **ubiquity** and **diversity** (pp. 150–155), or making no quantitative statements at all. Not only is this an unnecessarily pessimistic approach, it does not actually escape the problems of preservation (Kadane 1988: 207). Others have taken a middle ground by placing more confidence on ordinal data, such as saying that deer were more common than goats in an assemblage, but not by how much.

Archaeologists have devoted a lot of thought to ways of quantifying things, and careful selection of a quantitative measure in the context of a particular research problem can lead to useful results. The first step is to determine whether that research problem requires us to know the absolute number of whole entities like animals or pots, or just their relative abundances. Alternatively, it might call for estimates of amounts of food or raw materials, or just to infer whether one site had more artifacts of a particular type than another. It is important to realize that some measures are only simple ways of counting (enumerating) items *in* a particular sample (description), while others are indirect measurements or estimates of population characteristics *based on* a sample (see sampling in Chap. 6).

Most of the archaeological literature on quantification has focused on faunal remains, while some has focused on pottery or plant remains, and less than you would think on lithics. This chapter will employ examples from various domains, since most of the quantification problems are shared even though some categories of material culture illustrate certain problems more clearly than others.

There are many commonly-used measures for describing archaeological samples, and a few that are indirect measures or estimators of population parameters. This chapter will review the most common ones to describe how they work and their strengths and weaknesses, beginning with simple ones that describe the contents of a sample of the fossil assemblage (see p. 242), and working up to ones that archaeologists use in an attempt to infer characteristics of some population that was once living or in use (**life assemblage**), or that died (**death assemblage**), or that was deposited in a site (**deposited assemblage**). The chapter will end with a few non-abundance measures, including ubiquity, diversity, and fractal dimension.

7.1 Assessing Abundance in Samples

The sections that follow introduce a number of competing measures and the quantification problem they are meant to address. To be precise about the factors that affect their value, or how to calculate them, there will be mathematical expressions that use the following terms:

- N The number of whole entities (e.g., animals, pots, or tools) that contributed to the sample
- s The number of elements (e.g., bones) in a whole entity
- r The recovery rate, or probability that any element will be deposited, survive and become included in the sample (thus it is really the product of several probabilities)
- f The fragmentation rate, or average number of pieces into which whole elements tend to break

In almost every case, N is unknown and we can only estimate r and f and their errors, but we can sometimes know s quite accurately.

The discussion that follows these expressions will explain them in ordinary language and provide an example. Keep in mind that the general problem all these measures address is validity: to what extent does the indirect measure we use reflect the values of the thing we really want to measure?

7.1.1 NISP

If we only want to describe what is in our sample, the most obvious way to quantify items is simply to count them. Archaeologists typically call the absolute number of identifiable potsherds, seeds, flakes, bones, and bone fragments NISP, or "Number of Identifiable Specimens" (Chaplin 1971: 64–67; Grayson 1978). The less commonly used NS ("Number of Specimens"), by contrast, is just the number of items, regardless of whether they are identifiable or not.

While we really do not need an equation for NISP, as it is a simple count, we can use a mathematical expression for the factors that lead to its value (Chase and Hagaman 1987; Ringrose 1993a):

$$\mathrm{NISP} = Nrsf$$

What this means is that the value of NISP depends in part on the number of whole entities (N) of animals, pots, or trees that contributed bone fragments, sherds or charcoal fragments. However, it also depends on the number of potentially countable items (e.g., bones, seeds or handles) in a whole animal, plant, or pot ($s \geq 1$), their fragmentation rate ($f \geq 1$), and recovery probability ($r < 1$).

For example, where we have a type of bone with high survivability in our particular site context, that occurs twice in the animal's body, and that tends on average to break into four identifiable pieces, if our sampling fraction of the site is 5% or 0.05, we would expect (assuming easy identification and perfect recovery within the excavated fraction) the number of fragments of those particular bones' fragments in our sample to be $N{*}.05{*}2{*}4 = 0.4\ N$ so that, if there were 10 animals (N = 10) represented in a random sample, we would expect a NISP of 4. For zooarchaeological and some other kinds of samples, we would have multiple types of elements, each with different values of r, s and f, and, when we make an effort to avoid counting the same bone or pot more than once, it is really an attempt to soften the effect of f.

Simple as NISP may appear, it has important limitations. Thanks to those pesky r, s and f, the NISP of sherds does not tell us anything directly about the number of pots, and the NISP of bone fragments does not directly represent the number of bones, let alone the number of animals, that contributed to the sample. NISP is also sensitive to our decisions about what is "identifiable" and to whether we count conjoinable fragments as one or two observations. Typically, these decisions depend on both size and the presence of features, or "landmarks," that allow us to assign fragments correctly to classes (Table 7.1). Even for some unfragmented remains, such as whole seeds or beads, NISP would provide the basis for only indirect measurement of more interesting quantities, such as amounts of food, necklaces, or wealth.

Some have suggested that NISP lacks the quality of independence — the property that the presence of one bone fragment has no influence on the probability that some other bone fragment will occur — because it involves the clearly incorrect assumption that every identifiable specimen comes from a different animal or other whole entity. When we count, say, 10 bone fragments to yield NISP = 10, it is possible that these came from 10 different animals, 10 different bones from the same or different animals, or are 10 fragments from the same kind of bone in one or several animals.

Table 7.1 Examples of "landmarks" that can be used to identify an element and reduce the risk that any element is counted more than once, and example ranges of fragmentation (f) in which the element is countable (before the element is so fragmented that it is no longer identifiable). For most artifact types, usually only one or two of the f fragments exhibits the landmark

Element	Example landmark	Range of f
Flake or blade	Platform	1–6
Vessel	Rim	1–50
Tibia	Lateral condyle	1–15
Dicot seed	Hilum	1–4
Tobacco pipe	Heel	1–15

However, there is no need to assume that the fragments come from distinct individuals if we are only interested in the *relative* abundance of taxa. Which exact animals were the source of the specimens makes no difference as long as the sample is representative of the population of bones and fragments, all animals have the same number of bones, all of which are equally likely to survive burial and end up in our samples, and all break into the same number of fragments, on average. The same is true of quantification by mass and area (see below). For that matter, any sampling with replacement (see Chap. 6) involves counting some items more than once. The key to preserving independence is not ensuring that nothing is counted more than once, but to be confident that the inclusion of one observation has no effect on the probability of including some other observation. That is a problem for most of our measures since archaeological items tend to be clustered, and it is possible that inclusion of a particular bone fragment in a sample might increase the probability that the sample will also include a fragment to which the first fragment used to be attached.

Furthermore, as f, s and r in the equation above emphasize, it is not realistic to think that all animals or other entities are identical in their number of skeletal elements or the probability of bone survival or fragmentation. So, if we want to use NISP as a proxy for enumerating taxa of greater interest, such as whole animals, we need to consider the probable effects of these confounding variables.

Archaeologists have paid considerable attention to the effect of f, finding that it is not as simple as the equation implies. As fragmentation increases, this at first tends to increase NISP, as identifiable pieces break into multiple identifiable pieces, but then some fragments start to become too small to be identifiable, so that NISP decreases again (Fig. 7.1), potentially reaching the point where there are no identifiable pieces. Fragments that belong to different classes in our classifications typically have substantially different but imperfectly known fragmentation rates, leading some classes to be over- or under-represented (Chaplin 1971: 65–66; Orton et al. 1993: 209; Ringrose 1993a). Zooarchaeologists have attempted to control for this by examining the density

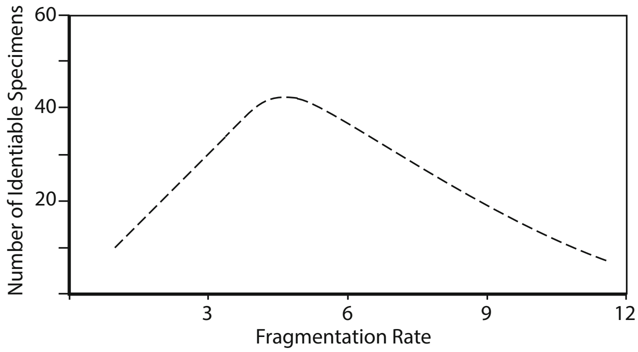


Fig. 7.1 Effect of fragmentation on the value of NISP, here for 10 complete specimens. At first fragmentation leads to higher values of NISP, but then many fragments become unidentifiable, so that NISP approaches 0 at high levels of fragmentation

and shape of different bones (see Chap. 15). This problem has received less attention for non-faunal materials but it is clear, for example, that heavy bases and handles of pots have lower fragmentation rates than vessel walls.

One problem has been virtually ignored: although we can often know *s* (at least for animals and some kinds of artifacts), and may be able to estimate *r* for the **preserved** or perhaps **deposited assemblage** (but generally not the **death assemblage** or **life assemblage**, see p. 242), it is technically impossible to know *f* (it has **fractal** properties, see below) unless we specify the minimum size of fragment that is countable.

On the other hand, the problems with fragmentation have an up-side. When we discover that different materials or materials in different contexts have different values of *f*, this is helpful for inferring the site-formation processes that deposited or disturbed those materials (Egloff 1973).

Annoying as fragmentation is, recovery probability is also problematic. It is actually a combination of probabilities: the probability that a particular specimen will be deposited at the site, the probability that it will survive in the site's deposits long enough to be discovered, and the probability that some archaeologist will find, save, and record it. We can control to some extent for the last ones, since they are sampling fractions that depend on archaeologists' research strategies (Chap. 6) and methods, but we usually have no basis even to guess the first one. We also have to keep in mind that an archaeologist's success at finding a large or obtrusive item is greater than for small items and ones that are easily mistaken for things like natural stones, a problem sometimes called visibility bias (see PPS sampling, p. 97).

Finally, even for animals that have the same number of bones, *s*, they may still differ in the number of *identifiable* and *distinguishable* bones (Lyman 2015). In that event, simply using the number of bones in whole animals as *s* can be misleading because, for example, we may be unable to tell the difference between two kinds of vertebrae from one taxon, or cannot identify the genus or species of a rib fragment.

After this rather pessimistic assessment, you might wonder why we bother with NISP. Despite its problems, it has several advantages. First, it is easy to calculate; you only need to keep a tally of specimens as you catalogue them. Second, ordinary NISP values are additive. That means that when you increase your sample size at a site or combine two samples, you only need to add new NISP values to the old ones or add two NISP values together (unless the analyst is not counting fragments that join to others). By contrast, some of the measures outlined below require you to start all over. Third, NISP may be better than at least some of the other measures when you want to compare proportions of fragmentary remains in two or more assemblages. Although it is certainly not unbiased, the factors *r* and *f* may be similar enough in some assemblages being compared for us to make reasonable judgments about their relative similarities and differences as long as the samples are sufficiently large, and include the same classes. Fourth, many of the problems with NISP are shared with more complex measures that are far harder to implement, making it the simpler choice. Measures based on NISP can serve as measures of fragmentation itself, and thus help us understand the taphonomy of particular sites or deposits.

7.1.2 The Shotwell Measure

As noted, one of the factors that distance NISP from *N* is the number of elements, *s*, in a whole entity. Shotwell (1955) introduced what he called "corrected number of specimens" as a simple correction for this factor for faunal samples, as *s* of animals is generally well known:

$$NI_s = \frac{NISP}{s}$$

For example, if we were to count clam shells with a NISP of 64, we would divide this by $s = 2$, since clams are bivalves (i.e., they have two shells). This yields a Shotwell index of $NI_s = 32$.

Simple as this may seem, it is not always obvious what the value of *s* should be. For example, do we count the proximal and distal parts of long bones as separate elements? Do we count elements that are almost never found? Should we group elements, like some vertebrae, that are so similar that we usually cannot tell them apart? How do we count bones that are fused together, or fragments that join? This calls for well-documented protocols. In the case of the simple example above, we might only count complete shells and fragments that included the umbo as a landmark (Table 7.1 and see p. 250), which then also accounts for *f* in NISP. Most analysts who use this approach use a modified value of s that reflects the set of elements that are potentially identifiable.

It is not always obvious how to adapt Shotwell's approach to non-faunal remains, however. For pottery, we cannot expect pots always to have the same number of separate parts, although some kinds of pots might always have two handles.

Holtzman's (1979) "frequency of elements" (FE) and Binford's (1978: 70) "MNI" are really adaptations of Shotwell's measure to individual elements, rather than whole animals, which is useful when we are interested in the selection of particular carcass portions for transport to a base camp and similar issues.

Palynologists are attempting a similar kind of correction when they correct for the effects of differential pollen production and transport on pollen abundances in sediments (Faegri and Iversen 1989: 3, 118–20, 141–46; Moore et al. 1991: 183–84). Because some species of plant produce far more pollen than others, and some pollen drops to the ground within a few meters of the plant while others travel hundreds of kilometers, the simple pollen abundances in a sample are not a good reflection of the abundances of the plants that produced the pollen. Their "representativity factor" (R) operates somewhat like Shotwell's use of s, but also incorporates aspects of r, and is based on the ratio of pollen density in modern surface samples to the abundance of the species in the vegetation found near the sampling location (Davis 1963, 1965). Different R values are needed for different plant communities in different regions. Because palynologists typically rely on relative abundances (percentages), this adds the complication that the percentage of one taxon will go up or down merely because of changes in other taxa. The R factor for these cases is,

$$R_{rel} = \frac{P_{surf}}{V}$$

where P_{surf} is the percentage of pollen within a particular catchment on the modern surface that belongs to a taxon, and V is the percentage of that taxon among the plants in the modern vegetation surrounding the catchment (Moore et al. 1991: 184). For example, we might find that 18% of the pollen that we capture in a pollen trap from the air in a particular location comes from oak trees, while only 10% of the trees in the vicinity are oak. That means the R factor is 18/10 or 1.8.

R values cannot easily be generalized to cases outside the context in which they were measured. As an alternative, palynologists often group taxa into super-categories of plants, like pine, whose high rates of pollen production cause them to be overrepresented in the pollen samples (type A), a middle group (type B), and underrepresented taxa that have low pollen productivity (type C). They then adjust the counts of each type by a correction factor based on typical pollen-production rates before calculating percentages with the aim of preventing group A from swamping the other taxa on the graphs (Faegri and Iversen 1989: 126–27).

7.1.3 Minimum Numbers (MNI, MNV)

Archaeologists have long sought a way to estimate the number of whole entities or Actual Number of Individuals (**ANI**, typically animals or pots) that would account for the fragmentary evidence in a sample, or at least to eliminate the lack of independence they perceived in the fact that NISP and Mass ignore the fact that some individuals "get counted" more than once.

The classic example of this is MNI (Minimum Number of Individuals), apparently first used explicitly in paleontological research on the Rancho La Brea Tar Pits (Stock 1929; Howard 1930) but widely found in zooarchaeology by the 1970s:

$$MNI = \max \sum s_i$$

What this means is that it is just the most abundant (maximum) of s elements: if we count how many of each skeletal element occurs in a sample, counting left and right elements separately, MNI is just the most numerous element of those that occur only once in each skeleton. If the sample contains 38 left humeri, for example, at least 38 animals must have contributed to the assemblage. If no other element numbers more than 38, then MNI = 38. This would seem effective at making sure that no animal is counted more than once.

In reality, things are more complicated, in part because the humeri themselves can be fragmentary. In that event, it is necessary to be more specific about what constitutes an "element" and the most common part might be the proximal end of the left humerus, for example. Analysts also vary in whether and how they account for sex, age, or size, in calculating MNI. For example, some may use the equation above to calculate MNI regardless of these factors, while others carefully compare left and right humeri to see if any, on the basis of their size or age, might be from the same animal. The latter will refine the count by adding specimens of the right humerus, when their size, sex, or age makes it impossible for them to have come from the same animals as the left humeri already counted towards the MNI.

Furthermore, it is not strictly true that no animal could be counted more than once, as this depends on how we aggregate the data. For example, it is possible that a gazelle's left distal humerus found in layer 6 at a site, where that was the most abundant gazelle element, was a residual bone (see

p. 323) that came from the same animal as a right proximal tibia in an earlier layer, 10, where right proximal tibiae were the most abundant elements. As different researchers may use slightly different protocols for dealing with fragmentation, uncertain identifications, size, sex and age differences, degree of contextual aggregation, and other aspects of the analysis (Grayson 1978), it is important for these protocols to be well-documented and defensible.

Extending this approach to quantification of pottery (**MNV** or Minimum Number of Vessels) requires adaptation to the fact that vessels, unlike animals, do not have consistent numbers of parts. Rather than count the various kinds of part (rim, handle, base, etc.) to find the most numerous one, the usual practice has been to concentrate on rims, or occasionally bases, in an attempt to discover the minimum number of vessels that could have contributed to the total circumference of all the rims. To do this, it is necessary to measure the proportion of a whole circle for which each rim accounts (see pp. 188) and then sum these. If they add up to the equivalent of 4.7 circles, for example, at least five whole vessels must have contributed to the sample. However, archaeologists typically also want to account for the fact that a rim with a diameter of 10 cm must have come from a different vessel than one with a rim diameter of 20 cm, so it is necessary to measure both diameter and circumference of each rim sherd, decide how much variation in rim diameter is likely to have occurred in each pot (e.g., ±2 cm or ±10%, since even wheel-thrown pots are not perfectly circular), and then lump the data into ordinal size categories. Summing the circumferences for each size, rounding up, and then summing the rounded figures for all sizes yields an MNV that is higher than would result from ignoring the size factor (see Table 7.2). In some cases, it also makes sense to use fabric type (see Chap. 12), not just diameter, to determine that sherds could not have come from the same vessel.

There have even been attempts, not very common, to adapt MNI to plant remains, yielding a minimum number of whole plants that must have contributed to a sample assemblage (MacNeish 1967). It is more straightforward to adapt it to stone tools, with the simplest measure of **MNT** (Minimum Number of Tools) being just the sum of complete tools and the greatest of the number of proximal or distal fragments of tools (Mayer-Oakes and Portnoy 1993; Shott 2000), most often adjusted upwards to account for material differences.

MNI, MNV and MNT seem attractive to archaeologists because they give the impression that we are quantifying whole animals, plants, or pots instead of just their fragments, or at least are avoiding double-counting of individuals. Indeed, these are useful measures in some instances, such as a bone bed at a well-preserved and fully excavated catastrophic kill site, or the pottery smashed on the floor of a building when the roof collapsed. In those cases, the minimum numbers are probably fairly close to (although almost always slightly below) the actual number of individuals (ANI) that contributed to the assemblage.

Table 7.2 Example of calculation of rim MNV for a sample of circular vessels of a particular type, with a variety of rim diameters distinguishable within 2 cm, and ignoring fabric type. Total circumferences (measured as percentage of whole circles) less than 100% round up to MNV = 1

Diameter (cm)	% Circumference	Total % circumference	MNV
10–12	37		
10–12	43		
10–12	28	108	2
13–15	29		
13–15	15	44	1
16–18	12	12	1
19–21	9		
19–21	22		
19–21	18	49	1
22–24	24	24	1
Total	237	237	6
n	10		

However, most archaeological samples are not like that, with the result that MNI, MNV and MNT yield biased estimates of ANI in the populations that contributed to assemblages. Generally, these are substantial underestimates. MNI has several unfortunate characteristics when applied to fairly typical archaeological samples (Grayson 1981; 1984; Lyman 2018; Marshall and Pilgram 1993; Orton 1993; Plug and Plug 1990):

- MNI overrepresents rare taxa relative to common taxa because even one fragment has to round up to 1 whole animal or pot
- Ratios and proportions of taxa based on minimum numbers are biased
- The ratios of minimum numbers are also sensitive to sample size; the proportions of taxa in a small sample are not the same as in a large sample from the same population (Grayson 1981)
- MNI, MNV, and MNT are sensitive to level of aggregation: the most abundant element at one level of aggregation is usually not the same as that for another level. MNI for a whole site or stratigraphic level yields a different (lower) result than the sum of MNI for individual contexts or sample units within the site or level
- For the same reason, the results are not additive; any increase in sample size requires recalculating MNI from scratch, although this may not be too difficult in projects that have prepared for that eventuality in structuring a database (see Chap. 4).

Generally, MNI and similar measures work best in the situation of large sampling fractions and well-preserved, catastrophic deposits in which MNI possibly approaches ANI. If used to calculate the proportions of taxa in a sampled assemblage, let alone in a life or deposited assemblage, those proportions will be biased to an unknown degree, leading some researchers to suggest that it provides ordinal-scale indications at best (Lyman 2018).

7.1.4 Mass

The oldest alternative to counting items or their fragments is to "weigh" them. In practice, the so-called "weight method" is to use the items' collective mass, as measured in grams (weight is measured in Newtons), in order to calculate relative abundances (proportions):

$$Mass_x = \sum Mass_{i_x}$$

This just means that the total mass of type x is the sum of the masses for all i items and fragments of type x, as measured in grams. However, we can also express this in terms of the factors that affect its value:

$$Mass = Nbr,$$

where b is the average mass of all the elements (e.g., bones) in a whole individual animal, or the average mass of wood in a tree or pot of a particular taxon, in the case of charcoal and pottery. If the recovery probability, r, was much the same for all the woody parts of the taxa of trees represented in a sample, for example, charcoal mass would be proportional to N, the number of trees (or, perhaps more usefully, volume or mass of wood fuel) that contributed to the charcoal assemblage.

The rationale for measuring the mass of pottery, bone fragments, charcoal, and other archaeological materials is that it is much less sensitive to differences in fragmentation (Solheim 1960), as long as you are careful to specify the size of the smallest fragments you include in the measurement or any other criteria that may have caused you to exclude something (see fractal dimension, pp. 121–122). With the exception of materials that are so badly preserved that identification is uncertain or impossible, it does not matter whether a pot or block of charred wood is in one piece or one hundred; its mass should be the same except for the small proportion of material that broke into pieces too small to be recognizable or recoverable. Fortunately, that small proportion would have contributed very little to the total mass, so any bias due to that is also small.

Mass is probably the most common way to measure the abundance of wood charcoal from archaeological deposits. In cases where we have reason to believe that the charcoal comes from the use of wood as fuel, mass of charcoal also has fairly high validity as an indirect measure of fuel use, even though it is not a perfect correlation, as fire may have consumed some kinds of fuel more completely than others. In the extreme case, surviving charcoal could even belong to a taxon that saw less use than one that has rarely survived (difference in r). As with NISP, the total mass of some taxon of wood charcoal could have come from one or several trees, but this does not matter (does not cause a problem with independence) if we are only interested in the proportions of various taxa that contributed to the assemblage.

Zooarchaeologists have also sometimes used mass as a measure that might be more valid than some of the alternatives for helping us understand the amount of meat that a sample represents (Barrett 1993; Chaplin 1971: 67–69). Advocates of this measure for faunal samples cite the following perceived advantages (Lyman 2008: 95):

- We can easily merge measures of bone mass into more general categories (e.g., family or even "deer-sized mammal" instead of genus)
- Total bone mass of each taxon is little influenced by fragmentation
- Bone mass has relatively high validity as an indirect measure of usable meat

However, the first of these is not unique to mass measures, and could even be applied to NISP, while fragmentation does still affect the identifiability of bones.

The most serious threat to the validity of this approach, however, has to do with the relationship of bone mass to usable meat. There is an allometric relationship between bone mass and carcass mass or muscle mass that varies in animals of different sexes and ages (see below, and Chap. 15). Large animals have greater bone mass than small animals and the ratio of bone to soft tissue in animals also varies with sex and non-linearly with age. However, as those problems pertain to the use of mass as an indirect measure of something else (mass of meat or number of living animals), they do not prevent us from using mass as a simple description of abundance in a sample.

Archaeological ceramic analysts have often been reluctant to use mass to quantify sherds because of their perception that, even though the mass of sherds is very close to the mass of whole vessels (Chase 1985: 215), there is a risk of overrepresenting large, thick-walled vessels over thin-walled fine wares. That bias would negatively affect our use of mass as an indirect measure of the number of vessels if we do not calibrate to the expected mass of whole vessels. However, it could arguably be advantageous if we are more interested in the volume of food or other materials that may have been stored in those vessels (see also Rodriguez and Hastorf 2015). There is a strong argument that mass provides unbiased estimates of the proportions of pottery types in an assemblage because fragmentation has little or no effect on the total mass of each type (Orton 2000: 52).

Another problem with mass is that it is vulnerable to differential preservation. Post-depositional processes such as leaching, mineralization and corrosion can remove mass from or add it to buried materials, something particularly noticeable in the case of metal artifacts and bones.

7.1.5 Area Measurements

A less common way to measure abundance is by surface area. This has been attempted for pottery, but probably makes the greatest sense in the case of window glass, whose function is directly related to area, while its mass depends on both area and thickness (Baxter and Cool 1990).

Hulthén (1974) recommends measurement by area to quantify pottery when we want to account for the sizes of the pots that contributed sherds.

$$A_x = \sum a_i$$

A_x is the total surface area of all the pottery of a particular type x, and a_i is the area of each sherd included in the sum. Because it is (or at least was at the time) tedious to measure sherds' surface areas, however, she recommends an indirect measurement based on the sherds' mass, mean thickness, and density. Similarly, Byrd and Owens (1997) recommend what they call "Effective Area," based on the aperture size of the screen that catches each sherd in a stack of nested sieves (see screen diameter, Fig. 17.5).

The area approach takes differential preservation into account, but not the other factors (such as varying sizes of whole pots) that intervene when we want to make population inferences on the basis of a sample. Some lithics analysts have also used either nested screens or size templates to short lithic artifacts roughly by area (see p. 223).

Baxter and Cool (1990) suggest a practical method for estimating the total area of window glass that contributed to an assemblage of broken window glass. This involves predicting area on the basis of mass and thickness in a sample from a larger population, using a regression analysis to estimate the relationship between area, mass and thickness. Their method is effective for estimating the total area of glass of each type that is in a particular sample assemblage, on the basis of a still smaller subsample, thus providing considerable time savings when the glass assemblage is large. However, they do not predict the total area of all the glass that contributed to the sample assemblage (i.e., the "life assemblage" of windows in use at a site at a particular moment in the past).

7.1.6 Ratios

While archaeologists are sometimes interested in the actual number of animals, pots, or other entities that went into a deposit, most of the time they are more interested in relative abundances, that is, the proportions or percentages of taxa. However, archaeologists have to base these ratios on one of the other measures already mentioned, typically NISP or mass. So, we might have proportions of different taxa of seeds,

$$P_x = \frac{\sum x}{\sum x + \sum (not\ x)}$$

This means the proportion of seeds of type x is just the sum of the counts of type x seeds divided by the sum of counts of all seeds, both of type x and not type x. So, if we count a total of 110 grape seeds and 890 other kinds of seeds in the sample elements from a site, the proportion of grape is 110/1000 or 0.11 (11%), with appropriate statistical errors (see cluster sampling, p. 98).

A classic example is the palynologist's pollen diagram (Fig. 7.2), which shows fluctuations in the relative abundance of pollen of different taxa over depth in a sediment column. One problem with relative abundances is that, when the proportion of one taxon goes up, others must go down, as the whole must add up to 1.0 or 100%. In other words, the various relative abundances are not independent of one another and, even if one taxon had an unchanging abundance over the whole time interval that the sediment column represents, it would still appear to fluctuate because of the behavior of the other taxa. In addition, whenever one taxon has relative abundance close to 100%, the fluctuations are virtually undetectable because the most abundant taxon overwhelms everything else (Faegri and Iversen 1989: 123–26; Moore et al. 1991: 170–74). Ratios are not additive either, although recalculating the ratios is not difficult.

However, some archaeologists have favored ratios as a way to mitigate differential preservation or different uses across contexts (Miller 1988: 75–83). A common version of this is to quantify charred seeds as a ratio to charcoal or charred nutshell. This involves the assumption that the amount of charcoal is a reflection of daily, domestic fuel use (not catastrophic burning of structures) and that the inclusion of nutshell in domestic fires is essentially random. The units may vary, and it is common for the ratio to be (counts of seeds):(mass of charcoal in grams) or (counts of seeds)/(mass of charcoal in grams). Some archaeobotanists would argue that this helps them control for preservation, as unusually high counts of some kind of seeds relative to charcoal would likely signal intensive use or a different kind of food preparation. However, this approach is

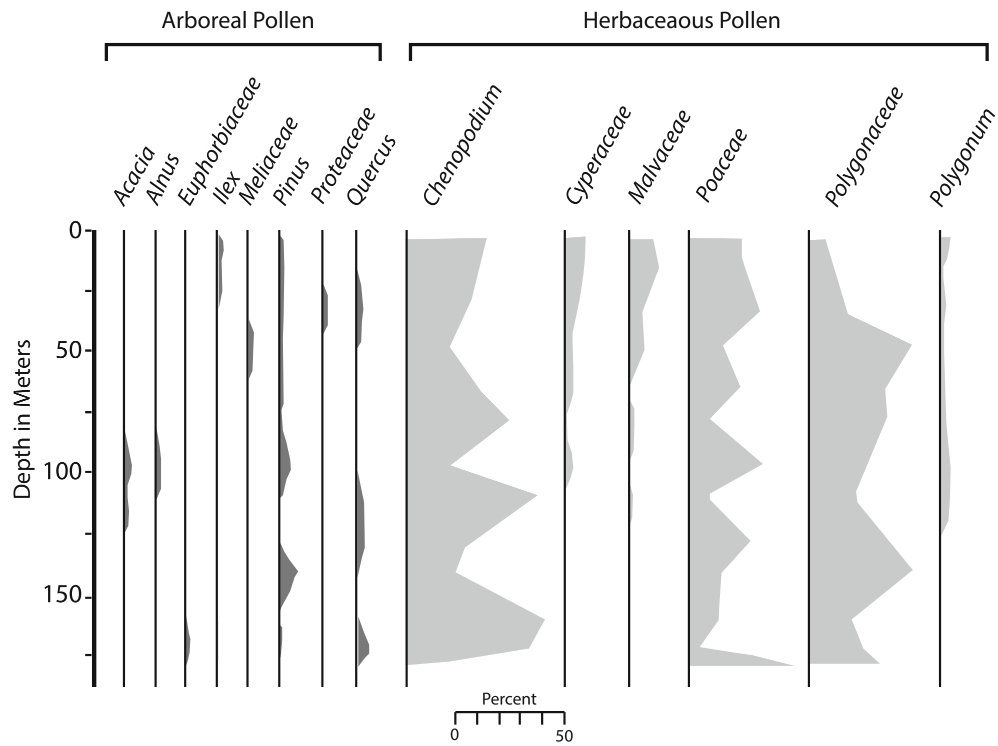


Fig. 7.2 Conjectural pollen diagram with various taxa of arboreal and herbaceous plants. Note how major peaks in one taxon tend to correspond with troughs in the spectra of other taxa. (See also Fig. 16.14)

vulnerable to the assumption that "daily" use of charcoal is a rather stable baseline; as with the pollen diagrams just mentioned, charcoal values could easily fluctuate themselves, potentially creating spurious peaks in the ratios of seed taxa. Users of this ratio should also be careful to consider possible changes, over time, in the types of fuel or the types of fuel-consuming activities. For example, introduction of the use of dung as fuel would likely cause a decline in charcoal abundance that could give the inaccurate impression of an increase in seeds. Furthermore, ratios do not escape problems of preservation; uncritical use of ratios depends on the assumption that the remains in the numerator and the denominator have similar probabilities of preservation (Kadane 1988: 212; see Orton 2000: 65–66). Probably it is best to use the ratio approach only in cases, like the seeds:charcoal example, where there is a meaningful relationship between the numerator and denominator.

7.1.7 Densities

To avoid ratios' problem of lack of independence, some archaeologists and archaeobotanists have favored densities (e.g., Miller 1988: 73–74), such as the number of artifacts per square meter of space or the number of seeds or mass of charcoal per cubic meter of sediment excavated, even though these are technically just another kind of ratio. An easy way to accomplish this is to count the number of each taxon by NISP or the Shotwell index, and divide by the sample area in m^2 or the volume of sediment in m^3, e.g.:

$$Density = \frac{NISP}{Area}$$

For example, we might find or estimate that there were 130 charred barley seeds in sediments that came from inferred storage silos whose total volume of sediment was 10 m^3, while 26 barley seeds came from hearths with a total sediment volume of 4 m^3. The density of barley in the silos would then be $130/10 = 13$ seeds/m^3, while that in hearths is $26/4 = 7.5$ seeds/m^3. We would need to consider the measurement errors on the volumes and statistical errors on the counts to see if this is a significant difference (see cluster sampling, p. 98, and Chap. 8).

Density measures allow each taxon to vary independently of the others, but there are other contributors to density that may complicate things. Variation in density can be due, not only to differences in the abundance of the items of interest, but also to sedimentation rates, erosion rates, stoniness of the sediment, and other factors (Chap. 17). Furthermore, density measures do not escape the problems inherent in the counting method, whether NISP, mass, or other measures, on which they are based (Table 7.3).

Table 7.3 Comparison of the features of different quantification measures

Measure	Ease	Additive	Immune to aggregation	Sensitivity to *f*
NISP	High	Y	Y	High
Shotwell	High	Y	Y	Medium
WAE	Medium	Y	Y	Medium
MNI or MNV	Medium	N	N	Medium
Mass	High	Y	Y	Low
Area	Low	Y	N	Low
Ratio	High	N	Y	Low?
Density	Medium	N	Y	High
Krantz	Medium	N	Unclear	Medium
Peterson	Medium	N	Unclear	Medium
EVE/ETE	Medium	Y	Y	Low
Ratio-of-ratios	High	Y	Y	Low

All are sensitive to degree of fragmentation (f), but "low" sensitivity means that this effect is minimized

7.2 Estimating Abundances in Populations

All the above types of quantification work to characterize samples. Even though some of them involve attempts to account for one or more of the factors that intervene between population and sample (or between life assemblage or deposited assemblage and recovered assemblage), they are not really estimates of population abundances in the sense of animals, pots or other whole entities that once lived and died or were used and discarded at or near an archaeological site. Even when the raw numbers come from well conceived cluster samples (see p. 98), those measures suffer from their inability to compensate well for differential preservation and recovery. The measures we will consider next are ones explicitly meant to account for these factors, at least in principle.

7.2.1 The Krantz Estimator

Krantz (1968) made the first attempt to provide an estimator for the number of individuals, N, in the population (death assemblage for animals) on the basis of a sample. Thus it accounts for all individuals that contributed to a deposit at a site, whether or not they occur in the sample at hand. However, Krantz's method is only applicable to paired elements, such as are found in vertebrate skeletons.

You may be thinking, "how can you possibly account for cases that are not in the sample?" However, we do this all the time on the basis of sampling theory (Chap. 6). In this case, Krantz cleverly takes advantage of the relationship between left and right elements in skeletons. For any skeletal element that occurs in pairs, the number of pairs, the number of left elements, and the number of right elements in the living population and the death assemblage would be equal ($P = L = R$). In any sample of that population in which we also found $p = l = r$, there would be no reason to assume that any bones had been lost, and it would be reasonable to conclude that the sample included all the bones of that element of all the animals that originally contributed to the assemblage. In other words, there was a 100% sampling fraction for that element (perfect recovery) and we could conclude that $l = r = p = N$.

In most archaeological samples, by contrast, many bones have gone missing through various site-formation processes, and others are missing because our sampling fraction is well below 100%. The fact that, in a typical sample, the numbers of left and right elements are not the same is itself proof that some are missing, as L and R in the life population must have been equal. Use of MNI involves the improbable assumption, despite the evidence that many bones are missing, that bones must have come from the same animal unless there is evidence to the contrary (Lie 1980). Krantz instead takes the evidence for missing bones seriously.

If we assume that the processes that removed bones of a particular element and taxon acted randomly, all elements of that type have an equal probability of *not* occurring in the sample. In that case, we expect some animals to be represented only by the left element, others by the right element, some by both left and right, and some by neither left nor right. Thus, it is highly likely that many individuals are not represented at all.

The Krantz estimator makes use of this information, while making the additional assumption that the samples of left and right elements are independent. If, for example, the attrition processes, including archaeological sampling, removed 80% of the left *tibiae* from the sample (i.e., $l/N = 0.2$), where l is the number of left *tibiae* remaining after attrition, we would also expect that 80% of these left *tibiae* should have no matching right *tibiae*. This means that $l/N = p/l$, where p is the number of pairs remaining in the sample. By rearranging terms, we get,

$$N = \frac{l^2}{p}$$

and, since the probabilities of left and right elements being removed to create the sample are equal, it also follows that,

$$N = \frac{r^2}{p}$$

Krantz takes these two separate estimates and averages them to provide an estimator, N_K:

$$N_K = \frac{l^2 + r^2}{2p}$$

So, in a sample of, say, deer femora, we might find 26 left and 39 right femora, of which 12 appeared to be paired (on the basis of size and symmetry). Putting these figures into the equation yields $(26^2 + 39^2)/(2 * 12) = (676 + 1521)/24 = 91.5$. Note that this is a much larger number than the counts of these femora because the large difference between left and right and the relatively low number of pairs indicate that many bones are missing from the sample.

The Krantz estimator provides consistently higher estimates than the Peterson estimator except when l = r, when they provide the same result (Fieller and Turner 1982). Because it is very similar to the Peterson estimator in concept, we will leave discussion of its advantages and disadvantages to the end of the next section.

7.2.2 The Lincoln-Peterson Estimator

The estimator known variously as the Lincoln index or Peterson estimator, like the Krantz estimator, has the goal of estimating the number of individuals in a population (arguably the death assemblage) on the basis of a sample (Fieller and Turner 1982; Poplin 1976; Wild and Nichol 1983). It is modelled on the "sample-resample" or "capture-recapture" method found in ecology, wildlife management, and some kinds of archaeobotanical sampling (see pp. 285).

If they want to estimate how many bears are in a national park, for example, park managers do not round up all the bears and count them. Instead, game wardens or wildlife biologists attempt to sample the bears randomly, tagging each bear and then releasing it. Let us say they sample and tag 100 bears. After enough time has elapsed to allow the tagged bears to be "well mixed" into the population, the rangers sample the bears again. This time, they not only count the bears in their sample, but also count how many of them have already been tagged. To keep things simple, let's say the second sample also has 100 bears, of which 20 are tagged. Assuming the second sample is also a random sample of the bear population, and 20% of them are tagged, we can conclude that one-fifth of the bears in the population are tagged, within the usual sampling errors. And since we know that the number of tagged bears is 100, that means that the total number of bears in the population is $100/0.2 = 5 \times 100 = 500$ bears. We can express this algebraically as,

$$N = \frac{n_1 n_2}{p}$$

where n_1 is the number of bears in the first sample, n_2 is the number in the second sample, and p the number of tagged bears in the second sample (i.e., number of bears in both samples).

Although archaeologists certainly cannot go back in time to tag anything, the Peterson estimator works by analogy to this method. In zooarchaeology, where the method has the clearest application, the left elements in a sample are analogous to the first sample, the right elements are analogous to the second sample, and the number of paired elements is analogous to the bears caught in both samples. If we can assume that the lefts and rights are well "mixed," and that we can confidently recognize pairs, then we define the Peterson estimator (N_P) as,

$$N_P = \frac{lr}{p}$$

This gives a somewhat lower estimate than N_K except when $l = r$. Using the same numbers as in the Krantz example, we would have $26*39/12 = 84.5$.

The sample-resample structure on which the Peterson estimator is based is also useful in archeobotanical and microrefuse studies that require analysts to count or estimate potentially hundreds of tiny pollen grains, seeds, or microscopic fragments of flint, shell or pottery in large volumes of sediment. Rather than count them all, which is very time-consuming, analysts can add a known number of some "exotic" particles, such as *lycopodium* spores added to pollen samples or tiny beads added to microrefuse volumes, and thoroughly mix the sediment prior to subsampling it. These added particles are then analogous to the tagged bears in the wildlife-management scenario above; because we know there are, say 100 beads in the volume of sediment, if we count 10 beads in a sample of it, we know our sample contains about 10% of the population and can then multiply our counts by 10 to get reasonable estimates of the total in the volume. We can also calculate standard errors and confidence limits for these estimates.

The main advantages of the Krantz and Peterson estimators are that they provide estimates of a population parameter, rather than just describing the sample, and have a probability distribution and confidence limits. We can calculate the confidence interval for the Peterson estimator by modelling it with the hypergeometric distribution, which is available in many statistical software packages.

However, both the Krantz and Peterson estimators pose some challenges in practical application and some of their assumptions (Table 7.3).

- They are only applicable to paired elements, which may occur in faunal assemblages, but are not easy to identify in other classes of archaeological materials. For a pottery type that consistently has two handles, for example, it is highly unlikely that we could distinguish a "left" and a "right" handle, or ones that were paired (except in substantially complete vessels). Even in faunal samples, paired and pairable elements can be infrequent or hard to identify.
- We cannot be certain that "matching" left and right elements, even for skeletal samples, actually came from the same individual, rather than different ones that were just very similar in size, unless the two are attached, as in mandibles or pelvises.
- When lefts and rights *are* attached in this way, it violates the assumption of independent random sampling — attached elements could hardly be "well mixed" — and thus lead to underestimating N (i.e., bias).
- It is conceivable that butchering practices favored left or right elements in faunal samples (Binford 1978: 70), so that the inclusion of lefts and rights in the sample is not random, which would be another source of bias.
- As with MNI, these estimators are not really additive because, upon increasing a sample size, it is necessary to check the new specimens against previously analyzed ones to see if any match to make pairs.
- They are also somewhat sensitive to fragmentation because fragmentation makes pairs more difficult to recognize.
- Whether or not they are vulnerable to degree of aggregation really depends on how confidently one can identify pairs. If there is a risk that lefts and rights from different contexts and different animals could be mistaken for a pair from a single animal when the contexts are combined, aggregation would be a problem.
- When the number of pairs, p, is very low, the standard error on the estimate of N is extremely large.
- When different elements give different estimates of N, it is not clear how to proceed, and this would seem to contradict the claim to have estimated "the original killed population size" (Fieller and Turner 1982: 54) rather than the fossil or deposited assemblage size (Ringrose 1993a). However, Fieller and Turner (1982: 55) note that this disagreement among elements in their estimates could be a very useful indication of reliability of the measure or "reveal patterns of differential deposition."

Overall, when the number of killed individuals that contributed to a faunal assemblage is of interest, it could be useful to use Krantz or Peterson estimators for their ability to estimate a relevant population parameter with a measurable degree of error and confidence interval. This measure could also contribute to indirect measurement of meat quantities. However, it is less clear how to apply these outside zooarchaeology, and users should be very careful about how they match elements to identify pairs and be sensitive to possible violations of their assumptions.

7.2.3 Completeness Indices and Estimated Equivalents (EVE and ETE)

Egloff (1973) and Orton (1980; Orton and Tyers 1992) formulated a measure useful for relative abundances, based on the completeness of rims, although Egloff focuses on using the measure to provide "minimum" numbers represented in a sample, as in MNV. What Orton calls Estimated Vessel Equivalents (EVE) yields unbiased estimates of the ratios or proportions of different taxa among pottery in a population, but a similar approach is applicable to bones, stone tools, and other items for which it is possible to measure the completeness of specimens.

$$EVE_x = \sum c_{xi}$$

where EVE_x is the vessel equivalent for rims or bases of type x (usually subdivided by fabric group, diameter, and perhaps other factors), and c_{xi} is the proportion of a whole circumference of the i-th rim or base of type x. For example, one rim might preserve 18% of the circumference while another with the same diameter and fabric type preserves 12% of the circumference, which sum to 30% or an EVE of 0.30.

EVE's first steps are thus identical to Minimum Numbers (MNV) calculations, but it omits the step of rounding up fractional values (i.e., we do not round EVE of 0.30 to 1.0). Instead of basing calculations on whole vessels, we base ratios on fractional vessels (Table 7.4). It does not matter whether two rim sherds of the same diameter and fabric come from the same or different vessels, sums based on the preserved circumferences of those rims will provide unbiased estimates of the proportions of vessel types in the population that was sampled. EVE also has notable advantages over MNV in providing more realistic estimates of the originating population size (e.g., Felgate et al. 2013). However, Chase (1985) expresses doubts about the accuracy of the circumference measurements (see p. 247), which, after all, depend on the assumption that rims and bases are almost perfectly circular. Of course, it is possible to take the uncertainty in diameter and circumference measurements into account (see, accumulating errors, p. 11).

Corredor and Vidal's (2016) alternative to EVE uses a modified rim count that accounts for the average number of rim sherds that result from breaking different kinds of

Table 7.4 A hypothetical example of proportions of three vessel types whose diameters are assumed to vary within ±1.25 cm to demonstrate the difference in outcomes between MNV and EVE (bolded). MNV provides biased estimates that exaggerate the importance of the rarest type (A) and somewhat underestimate type B, while EVE provides unbiased estimates of the proportions

	Total % circumference		
Rim diameter (cm)	Type A	Type B	Type C
1–3.5	140	0	0
4–6.5	230	290	80
7–9.5	110	350	0
10–12.5	420	160	110
13–15.5	40	0	250
16–18.5	0	420	0
19–21.5	0	680	370
22–24.5	0	210	130
25–27.5	0	0	90
MNV	13	24	13
MNV%	**26**	**48**	**26**
EVE	9.4	21.1	10.3
EVE%	**23**	**52**	**25**
NISP	46	115	74

pottery, but determining this average would usually require at least a pilot sample of rim EVE measurements.

In adapting EVE to bones and stone tools, the simplest approach is to make an index (Estimated Tool Equivalents, or ETE) based on the presence or absence of certain "landmarks." For stone tools, for example, we could characterize incomplete tools as proximal, medial, or distal fragments, and count any fragment that includes only one of these zones as 1/3 of a tool, while one that has, for example most of the tool except its proximal end as 2/3 of a tool (Mayer-Oakes and Portnoy 1993; Shott 2000). However, this approach is biased if there is any possibility that broken tools have more than one medial fragment.

It is similarly possible to treat identifiable bone fragments as proportions of whole bones on the basis of landmarks or "diagnostic zones" on them, or bones containing a landmark as proportions of whole animals on the basis of *s* (Rackham 1986).

Measurement Errors in EVE: Calculating errors for EVE requires us to accumulate the errors on the individual sherd measurements. Diameter charts only allow you to estimate the diameter and circumferential proportion of sherds within a few centimeters and a few percent, although 3D scanning now makes more precise estimates possible. In fact, the two measures are correlated, so a small error in estimating the diameter will also lead to an error in the circumference.

Example

It is possible to accumulate the estimated errors as described in Chap. 1. For example, for eight rim sherds whose circumferences sum to an EVE of 0.95, we might have estimated measurement errors as in Table 7.5. By squaring the individual errors, summing the results, and then taking the square root of that sum, we arrive at an estimate of the total measurement error on the EVE of 0.95 as:

$$\sqrt{.02^2 + .02^2 + .03^2 + .02^2 + .03^2 + .01^2 + .02^2 + .01^2} = 0.06$$

So, we have 0.95 ± 0.06. In addition, we would expect some sampling error, presuming that the eight sherds are a sample from a larger population (see Chap. 6). Having estimated proportions of pottery types on the basis of EVE, and assuming a cluster sample, the standard error on the proportions would be based on the sum of the squared differences between the proportions for each cluster and the "grand proportion" for the whole sample (Drennan 2010: 243–248), all divided by the square root of the number of clusters (see p. 98).

Table 7.5 The proportions of circumferences for eight rims and their estimated measurement errors, as based on the precision of the measuring instrument or assessment of lab personnel's performance

Sherd No.	% Circumference	Measurement Error
1	5	2
2	10	2
3	20	3
4	15	2
5	15	3
6	12	1
7	8	2
8	10	1
EVE	.95	.06

EVE and similar measures that are based on completeness indices are advantageous in cases where we want to estimate the proportions of different taxa because they provide nearly unbiased estimates of those proportions, along with confidence limits. However, they are not like the Krantz or Peterson estimators in that they do not estimate the actual number of individuals in a population (such as a life or use assemblage at a moment in time), but only the relative contributions of taxa within a deposited or sampled assemblage. For example, EVE does not account for differential use-life among vessel types (one type may break frequently while another tends to last many years before it breaks). As seen in Table 7.2, it also has the advantages of being relatively simple, additive, relatively unaffected by aggregation level, and less sensitive to fragmentation (the completeness index accounts for fragmentation except in extreme cases of destruction).

As for drawbacks, EVE and similar measures are slightly more time-consuming to calculate than just counting NISP and, in the case of pottery, vulnerable to errors in measurements of diameter and preserved circumference. Notably, Chase (1985: 217) claims that "rim arc" (EVE) "is not a reliable estimator of number of whole vessels." However, EVE is actually intended to estimate the proportions of pottery types, *not* total number. Ironically, in the small sample of eight vessels in Chase's quasi-experiment, the total arc for each vessel was actually very close to 360° (or EVE = 1.0), indicating that it performed just as we would expect.

7.2.4 Quasi-Counts (PIE and TIE)

For the purposes of statistical tests that require counts, rather than proportions, Orton and Tyers (1993) conceived of a pseudo-count transformation that turns EVE into PIE (Pottery Information Equivalents), based on the ratio of the sum of squared products of EVE and number of fragments to the sum of squared EVE values,

$$PIE = \frac{(n-1) * \sum (n * EVE)^2}{n * \sum (EVE)^2}$$

where n is the number of rim or base sherds, whichever was used for EVE. Using data from the ten sherds in Table 7.2 as an example, we would have,

$$PIE = \frac{(10-1) * \left((3*1.08)^2 + (2*.44)^2 + .12^2 + (3*.49)^2 + .24^2\right)}{10 * \left(1.08^2 + .44^2 + .12^2 + .49^2 + .24^2\right)}$$
$$= \frac{9 * (10.5 + .77 + .014 + 2.16 + .058)}{10 * (1.17 + .19 + .014 + .24 + .058)} = \frac{121.5}{16.72} = 7.3$$

These act like counts of pots even though they are not. Shott (2000) went on to adapt this approach to stone tools, with ETE (Estimated Tool Equivalents, see p. 142), and Moreno-Garcia et al. (1996) adapt it to faunal assemblages. Unlike MNV and MNT, which will always underestimate ANI, if only by a little, PIE will sometimes overestimate and sometimes underestimate it (remember the MNV for this example was 6), making it relatively unbiased.

7.2.5 Total Minimum Animal Units (TMAU)

Despite the name, this is not a version of MNI or even MAU, in fact, it is not a minimum at all. Like EVE, its purpose is to estimate accurately the relative abundances of taxa in a population. To accomplish this, Chase and Hagaman (1987) begin with a measure they call Weighted Abundance of Elements (WAE), which is similar to the Shotwell measure, but with elements such as vertebrae often grouped because they are indistinguishable (thus requiring a modified value of s). Where differences in average preservation, $\bar{r}$, are small and N in the deposited assemblage is not too small, they claim that the *relative* abundances in a population (probably the deposited assemblage) should conform to the expression,

$$TMAU = \sum_{i=1}^{g} \frac{NISP_i}{s_i}$$

where we sum NISP over g element groups, and s_i is the number of elements in whole animals for the i-th group of elements. Keeping in mind that we defined NISP as $Nrsf$ earlier in the chapter, this controls for the number of s elements because the s values would cancel out.

What this expression means is that we divide the frequencies in each group of elements by the number of those elements in whole animals (s_i), and sum these modified frequencies over all the element groups. For example, suppose we can only identify fish vertebrae to the groups, precaudal, caudal, penultimate and urostyle. If there are 14 precaudal, 15 caudal, and 1 penultimate vertebrae, and 1 urostyle in whole fish of a particular taxon (see p. 314), this fish's TMAU, if we had a sample of 26 precaudal, 56 caudal, 2 penultimate vertebrae and 1 urostyle, would be:

$$TMAU = \frac{26}{14} + \frac{56}{15} + \frac{2}{1} + \frac{1}{1} = 1.86 + 3.73 + 2 + 1 = 8.6$$

However, because different species may have different numbers of groups, this will provide biased estimates of taxonomic proportions unless you limit the analysis to element groups that all the taxa have in common. Alternatively, you could correct for this by dividing the values by the value of g for each taxon, or "Relative Frequency":

$$RF = \frac{TMAU}{g}$$

This is also biased if some of the taxa have more r_i values than others, so it is preferable, where feasible, to restrict analysis to the element groups shared by all taxa.

TMAU, like EVE but for animals, is a good attempt at providing a method for yielding good estimates of the proportions of different animal taxa in a population. Its main weakness is its assumption that the values of r will be

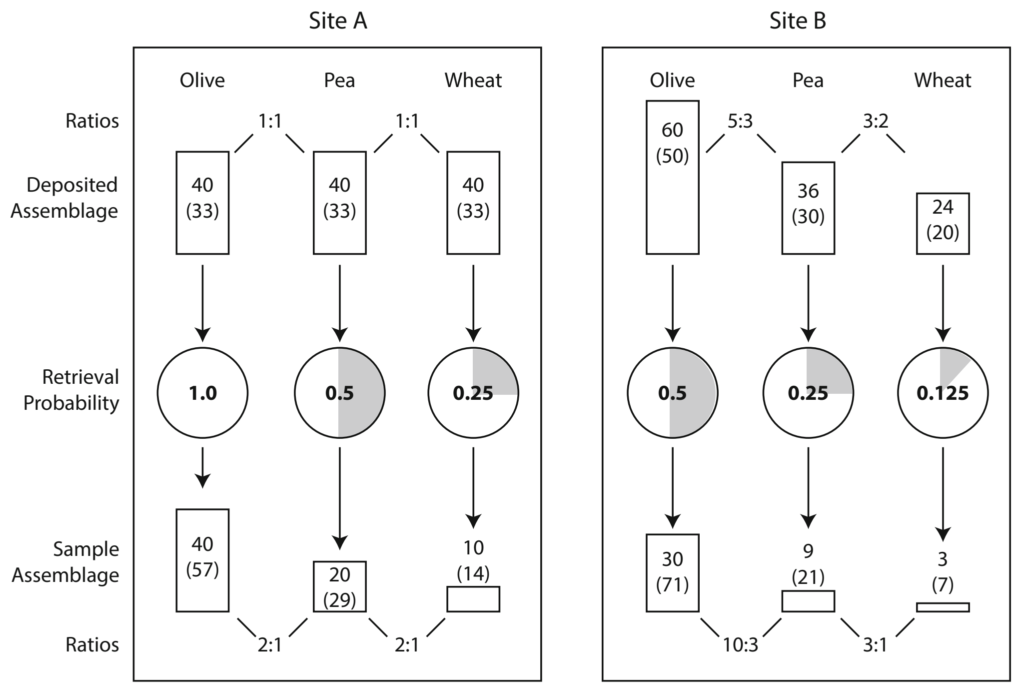


Fig. 7.3 Illustration of Orton's (2000: 65) ratio-of-ratios approach to estimating the difference in taxonomic ratios among sites with similar preservation conditions. In site A, the original ratios among three taxa are 1:1 and 1:1, while in site B the ratios are 5:3 and 3:2. After those assemblages undergo deterioration at rates that vary by taxon but are proportional at the two sites (site B is twice as destructive as site A), the ratios among taxa in the sample assemblages of the two sites preserve the original proportionality (e.g., the ratio of 1:1 to 5:3 is the same as 2:1 to 10:3)

similar for all element groups, as it is well known that *r* varies with bone density and other factors.

7.2.6 Ratios of Ratios

Orton (2000: 65–66) offers an approach that may be effective in cases where we do not need to know the actual abundances in the life, death, or deposited assemblages, but are really only interested in comparing sites or contexts in terms of these abundances. He notes that, when there are several taxa or different elements, each with different probabilities of preservation and retrieval, we can meaningfully compare assemblages among contexts if we are prepared to assume that the preservation probabilities in the contexts being compared are proportional. In other words, if the preservation probability of deer metatarsals is double that of, say, rabbit metatarsals at one site, it will also be double at another site, even if the actual preservation probabilities are different. When preservation "filters" the abundances of taxa at the sites being compared, the ratios of one taxon to another will be different in the final sample than they were in the deposited assemblage, but the ratios at one site as compared to the other will be proportional. If the ratio of deer to rabbits was twice as high in the deposited assemblage of site B as in site A, it would also be twice as high in the samples (Fig. 7.3).

7.3 Indirect Measures

7.3.1 Yield Estimates

Many archaeological research questions focus, not on the number of animals contributing to an assemblage, nor the proportions of plants or animals in that assemblage, but on the amount or proportion of kinds of food in ancient diets. This requires an indirect measure of food yield.

To accomplish this for animal foods (the most common application of this method), archaeologists have typically begun with one of the measures discussed above (usually MNI) and then transformed the data to account for the fact that different animals, or different bones in those animals, carry different amounts of meat, and potentially other useful food, such as bone marrow (Casteel 1974, 1978; Chaplin 1971; Lyman 1979, 2008: 84–113; Reitz et al. 1987; Smith 1975; Stahl 1982). Any transformation of data based on one of the measures above, of course, retains those measures' problems.

In addition to problems with the quantification measures that serve as the starting point for yield estimates, estimates of meat yield have their own sources of potential error. While we might be tempted, for example, to multiply the mass of identifiable red deer bones by the ratio of meat/bone mass in a modern deer's dressed carcass, the relationship is not linear and the ratio varies by the deer's age, sex, nutritional state, health, and possibly other factors. Consequently, simply pooling the data for animals of various sexes and ages and multiplying by a single ratio (such as an average) is not likely to provide a very realistic measure (i.e., it has low validity; Casteel 1978; Jackson 1989; Lyman 2008: 88–89). Basing the ratios on very small samples of modern animals is also unlikely to be representative of populations in the present, let alone the past (Ringrose 1993a: 148).

The study of the ratios of the metrics of body parts to one another is called **allometry**, and is the basis of this approach (Gould 1966; Lyman 2008: 94; Prange et al. 1979; Reitz et al. 1987; Reynolds 1977). Every individual living mammal has a particular relationship between, for example, its total mass and the length or cross-sectional area of its femur, or its stature and the length of its tibia, and these ratios change over the animal's lifetime (Jackson 1989; Noddle 1973). A plot of these changes with age on the x-axis will show a distinctive curve.

Some researchers have used multiple regressions of these relationships and applied them to the total bone mass of a species in an archaeological sample. One problem with this approach is that it is equivalent to assuming that all the bones came from a single, enormous, and very old animal, when in fact it probably came from a large number of much younger animals of various sizes and ages and quite different allometric ratios from those the faunal analyst used. Obviously, this provides biased estimates.

Obtaining more accurate results therefore requires sorting the bones and bone fragments into sex and age categories and then applying the appropriate ratio, with its error estimate, to each group (e.g., Barrett 1993). This limits the applicability of yield estimates to the subset of an assemblage that we can confidently assign to sex and age groups or, alternatively, we might content ourselves with a sort of composite or averaged ratio based on the sex and age distributions in that subset. Clearly, that would increase the size of the estimated errors on the yield estimates, but would allow use of more of the sample. As the proportional contributions of different taxa to diet constitute highly interesting information, these drawbacks are arguably acceptable. The mass allometric method seems particularly attractive for the analysis of fish remains, as long as the sample is effectively a random sample of all the vertebrae, because it is so easy to determine the age of fish vertebrae (see pp. 253).

However, applying a ratio of the meat:bone ratio in whole animals to individual bone elements when they come from quite different body parts is problematic (Lyman 2008: 103–104). We would not expect phalanges to carry as much meat as femora or humeri, for example, so multiplying the mass of such elements by a single ratio has problems of validity when it is possible that hunters, for example, carried only the most meat-rich portions back to their home base.

Particularly given the substantial variation in reported ratios of bone weight to live weight or meat weight (or the equivalent for plant foods or the contents of pots), it is also very important for researchers to report the errors on their estimates. As meat yields and similar values are indirect measures, we would expect their errors to be somewhat larger than the errors on either the quantification measure (e.g., MNI, N_K, or bone mass) or the meat:bone ratio (see "Propagation of errors," p. 11).

Somewhat similar to yield estimates are measures of the "utility" of different faunal elements or of differential use of body parts. We will return to those in Chap. 15.

7.4 Are Ratio-Scale Measures Justified?

The measures discussed above have varying strengths and weaknesses (Table 7.3), but all of them share some significant problems, particularly if we expect them to be indirect measures of inaccessible population parameters:

- The target population is often poorly defined
- The assemblages subject to direct measurement are not random samples of any population, with the possible exception of subsampling the "sample assemblage" (the population of items excavated and curated)
- The taphonomic processes implicated in the various transformations of "life assemblages" into "sample assemblages" and their impacts on the measures are imperfectly understood and are themselves often the subject of investigation
- The validity of the indirect measures is not well demonstrated
- Measures, like the Krantz and Peterson estimators, that purport to estimate population abundances, require assumptions that may be dubious, pose practical difficulties, or may yield different results for different elements

Given the very real difficulties inherent in quantification of archaeological abundances, some authors have suggested that we might better consider the results of such measures to be "ordinal at best" (Grayson 1984: 30; Lyman 2008, 2018; Wolverton et al. 2016). Rather than succumb to the temptation to compare assemblage abundances as though we have good ratio-scale parameter estimates, arguably we should use non-parametric statistics (Wolverton et al. 2016), which require fewer assumptions than methods like t-test, and graphical methods to interpret results, all the while keeping the sampling problems and various taphonomic and other confounding variables in mind.

On the other hand, this might be a little too pessimistic. In keeping with the general theme of this book, the alternative is to try to provide conservative estimates of the errors associated with any of our measures, which are easy to portray as "error bars" on graphs (see p. 85). Even ordinal measures have errors, but careful consideration of the errors on measures, whether ratio or ordinal, may allow us to distinguish between differences that matter and differences that are illusory. We should also be wary of ordinal rankings based on very small differences in NISP or MNI values, given the likely magnitude of the errors in these measures.

Finally, a good piece of advice is not to rely too heavily on any one measure, and always to keep in mind the purpose of the analysis and the nature, size, and sampling and taphonomic problems of the sample at hand (Morlan 1983; Kadane 1988; Popper 1988).

7.5 Non-abundance Measures

7.5.1 Fractal Dimension

As mentioned in Chap. 2, means do not capture the characteristics of a distribution of sizes, masses or densities because we have to make an arbitrary decision about how large a sherd, lithic, bone fragment, or even site has to be for it to be measured or counted. Theoretically, these things can be microscopically small, and changing the cut-off at the small end of our distribution has a substantial impact on the value of the mean and density, especially because small items tend to be particularly numerous. If we want to characterize the distribution as a whole with a single number (or a number and its associated error), we need a very different kind of measure.

One that is appropriate in this case depends on the interesting fact that the distributions of site sizes in a settlement system, of artifact sizes in a collection of broken pottery or lithics, and of artifact densities on a site, have fractal properties (Brown et al. 2005). Archaeologists have been aware for several decades, for example, that site sizes often follow Zipf's Law (Hodder and Orton 1976: 69), which describes the relationship between settlement sizes and their ranks as a power function (Fig. 7.4). What they have not always appreciated is that this is a fractal property (Cavanagh and Laxton 1994; Laxton and Cavanagh 1995). Fractal geometry describes the characteristics of many rough, complex, and messy phenomena in nature and culture. Despite what might seem like disorganization, fractals show self-similarity at different scales. What this means is that the pattern you see at a large scale looks much the same as the one you see at a small scale. Fragmentation, which is usually the bane of an archaeologist's life, has this characteristic; if you begin to study microdebitage — microscopic flint flakes from flintknapping (see p. 305) — you will likely find it surprising how similar micro-flakes are to large flakes, apart from their size, and the same is true of their size distributions at macro- and micro-scales.

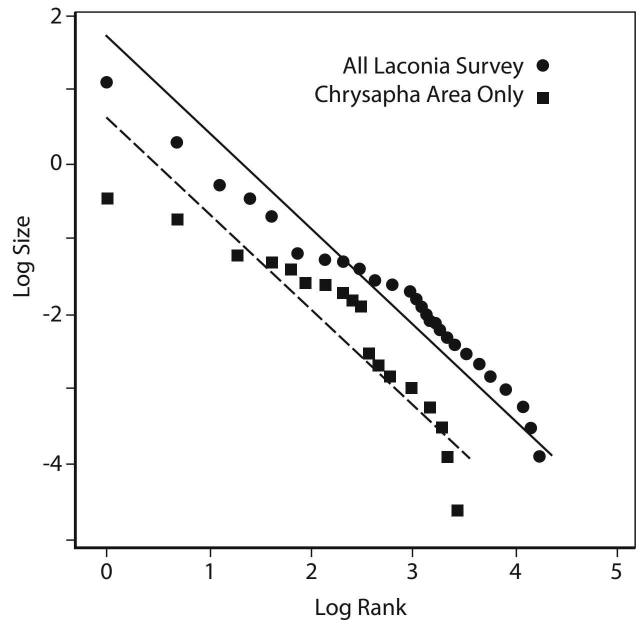


Fig. 7.4 Graph of ranked site sizes follows Zipf's Law, a power function that is a fractal property, here illustrated with data from the Laconia Survey, Greece, and a subset from the Chrysapha area within it (after Cavanagh 2009: 414). The fact that both the full data set and the subset have the same slope means that they have the same fractal dimension, demonstrating self-similarity at different scales

Although fractal geometry is much more than a single statistic, we can often characterize a fractal distribution with its fractal dimension (D), which is a non-integer dimension. Most of us are used to one-dimensional, two-dimensional and three-dimensional shapes, but fractals can have "weird" dimensions, such as 1.5 or 1.7. What this D value does is express the change in detail that results from a change in scale. D is the exponent in a power-law function,

$$f(x) = kx^D$$

where k is a constant. Another way of looking at D is as a ratio that describes the relationship of the number of self-similar parts (a) to the sizes of those parts (s):

$$D = \frac{\log a}{\log s}$$

We can characterize a distribution of sizes of fragments, such as sherds or flakes, with the same power law that describes fragmentation more generally (Turcotte 1986: 1922):

$$N(>s) \sim s^{-D}$$

This just means that the number of artifacts with a particular linear dimension greater than s, or N(>s), is approximately proportional to that size s raised to the power, $-D$.

Brown et al. (2005) demonstrate this with debris from flintknapping, which they passed through a set of nested screens and then counted how many flakes were on each screen to derive the logarithmic graph in Fig. 7.5. D is a ratio, and the negative slope of the regression line is our best estimate of its value. In this case, $D = 1.37$.

In the case of accidentally broken items, such as pottery and animal bones, it is likely that the fractal dimension tells us something about the fragmentation process itself. In fact, the value of D is directly related to the probability of fragmentation (Turcotte 1986: 1923–1924). Pots and bones with different densities (see density-mediated attrition, p. 243) and other physical properties break differently, while dropping, trampling, scavenger chewing, and hammering also affect the value of D, or may distort distributions in such a way that the distribution does not fit the power-law function at all.

We can study changes of scale in an archaeological spatial pattern by using the "box-counting" version of the D statistic. This involves imposing a grid on a map of site or artifact locations and counting how many sites or artifacts are in each "box" (N), then merging boxes to repeat this process many times, until everything is in just one box (this only works well when the study area is a rectangle). We expect there to be a relationship between N and the size of the boxes (s), which we can find by plotting the results on a logarithmic graph, much as with the flint debris example in Fig. 7.5. Again, the negative slope of the regression line on this graph provides an estimate of the ratio D.

In classic examples of rank-size analysis in archaeological settlement patterns, Laxton and Cavanagh (1995) found that we can relate the rank-size distributions to fractals by $D = -1/k$, where k is a constant that describes the scaling between settlements of different rank and, for Zipf's classic case, is $k = 1$. In their settlement systems in Laconia, Greece, D varied between 0.7 and 1.0.

7.5.2 Ubiquity

Some archaeologists use ubiquity (sometimes called "presence analysis") as a measure in contexts where one might have used an abundance measure, but ubiquity does not actually measure abundance. Instead, it is a measure of how "commonly" something occurs in a sample. More precisely, it is a measure of how often at least one of some item occurs in the sample elements (Diehl 2017; Hastorf 1988; Minnis 1981). This can have important implications, for example, in determining the likelihood that some archaeological search or sampling method will successfully detect certain kinds of artifacts, or will accurately characterize their abundance (Sullivan and Tolonen 1998).

The popularity of ubiquity in archaeology, especially among archaeobotanists, stems mainly from the impression that it is less sensitive to differential preservation than are competing measures. However, it most certainly does not escape that problem, as an ubiquity value, like all the values previously mentioned, depends on the probabilities of preservation and sampling. Even if it were to perform better than some other measures in this respect, it would only be at considerable cost of detail (Kadane 1988). Furthermore, it is not always clear what it is showing us.

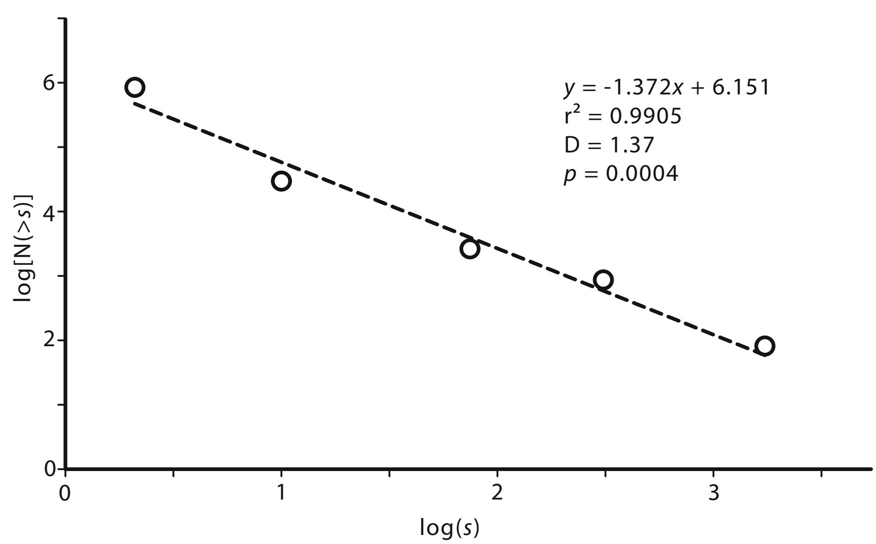


Fig. 7.5 Calculation of the fractal measure, D, for the sizes of flintknapping debris (Y. Salama, after Brown et al. 2005). The axes are natural logarithms, r is length of flakes in millimeters and N(>r) is the number of artifacts with lengths greater than r

Ubiquity is easy to measure, at least in those cases, quite common in archaeology, when we have a sample that consists of n equally sized spatial elements, such as squares in a grid or standard 10 L volumes of sediment for flotation (see p. 272). In that case, ubiquity is simply a proportion: the number of sample elements that contain at least one of some class of item (x), divided by the total number of elements in the sample (n):

$$Ubiquity = \frac{x}{n}$$

Example

For example, if we have 100 bags of sediment, each with a volume of 4 liters, and 48 of those bags contained at least one charred maize kernel, then ubiquity is just 48/100 or 0.48. At least one raspberry seed might occur in 80 of the bags, yielding an ubiquity of 0.8, but this does not indicate that raspberry is more abundant. There may have been only two or three raspberry seeds in each of those 80 bags, while many of the bags that had maize in them may have contained 20 or 30 maize kernels.

The appropriate statistical model for ubiquity is the binomial distribution (see Chap. 8). Its parameters are n (the number of "trials" or elements in the sample) and p, the probability of getting a "success" (e.g., at least one maize kernel or raspberry seed) on each trial, which we estimate from the proportion of successes in our sample. Methods for estimating confidence limits for ubiquity are the same as those for proportions, typically involving a normal approximation when n is large but binomial tables when n is very small (Wallis 2013).

Ubiquity has a number of limitations:

- It is only reliable when n is reasonably large and from a representative sample of a defined population
- It is very sensitive to both the size of the sampling elements and the density (e.g., number of seeds per liter) of the items
- It is meaningless when the elements are not comparable (i.e., unequal sediment volumes)
- It is of questionable value in cases where large differences in abundance or density are likely to be important
- Ubiquity might be useful for examining site-formation processes, but it is not clear how it relates to the "importance" of taxa. It most certainly does *not* indicate abundance

Some ubiquity values reported in the archaeological literature are based on sample sizes no more than 10, which makes them of dubious value. Keep in mind that, for $n = 3$, the only possible values of ubiquity are 0, 0.33, 0.67, or 1.0 while, for $n = 5$, ubiquity can only take 6 values, and for $n = 10$ it can only have 11 values. Consequently, many taxa are likely to be tied in terms of ubiquity, even if they differ a great deal in abundance. Furthermore, if we increase sample size only slightly when n is so small, it is rather likely that the ubiquity value will change quite a lot.

Deciding on a sample size is consequently very important (Diehl 2017). As usual, there are fixed-sample methods that could be used to predetermine the value of n that is likely to yield acceptable errors and confidence level (see p. 100), keeping in mind that the model for ubiquity is the binomial, not normal, distribution. However, there is a good case for using sequential sampling here if analysis cost is high (see p. 101). Having collected a large number of potential sediment volumes in the field, for example, you could randomly select and analyze some of them, plot the ubiquity values after each increase in sample size, and stop sampling when the ubiquity values level off and stabilize (Fig. 7.6).

But overall sample size is not the only issue. The size of the sample element and the density of items in those elements also have a significant effect. For example, if the sample elements each consist of 10 L of sediment, and densities are somewhat high, it is likely that, for many taxa, ubiquity will be 1.0 or only slightly lower. On the other hand, if the elements are only one liter or density is fairly low, ubiquity measures *even on exactly the same sediment* could be close to 0.1 or 0.2. Deciding on the size of sample elements is consequently just as important as sample size. Ideally, you want to have elements small enough that most taxa will have zero counts at least some of the time, but large enough that zeroes do not dominate the frequency distribution. Since we can model densities with the Poisson distribution, examining the distribution for values of λ based on a small pilot sample would be a good guide to appropriate volumes (see pp. 132–133). Keep in mind that the population that ubiquity is characterizing is not "the site;" it is the population of all potential sediment volumes of size x, and whether $x = 10$ L or 5 L or 1 L will make a big difference.

Ubiquity greatly reduces the amount of information available for analysis (Kadane 1988). If variations in abundance are likely to be important, it is unwarranted to forgo counts in favor of simple presence or absence of taxa.

The biggest problem with ubiquity results from the impression that it indicates abundance, as in Boyd's (1988) description of species as "very abundant" when he means to say "highly ubiquitous." It does not. High ubiquity really only tells us that a certain taxon is rather evenly spread through our population, instead of clustered, without saying whether it is abundant or not. Low ubiquity means that the

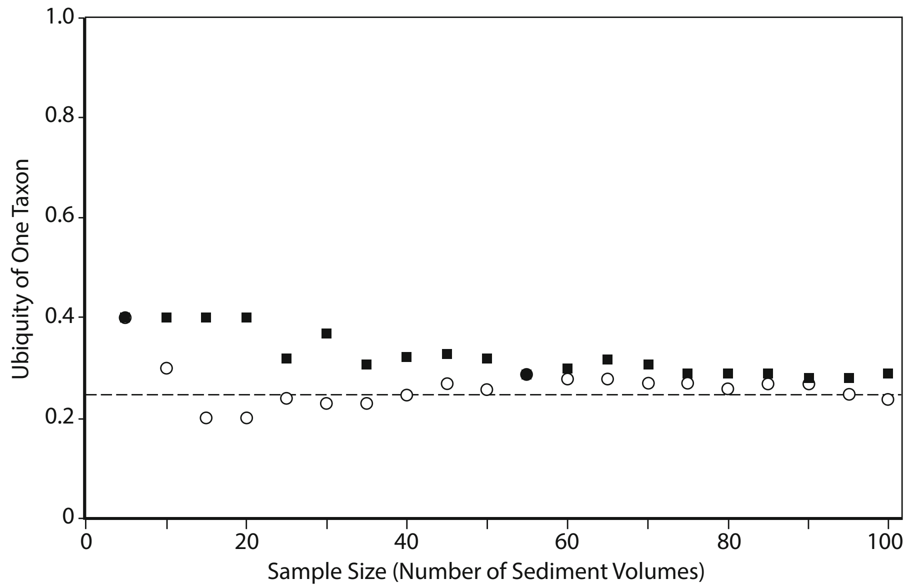


Fig. 7.6 Sequential sampling to determine how large a sample (n) is necessary to obtain a stable estimate of ubiquity. In both examples shown here, the probability that any particular sample element contained at least one item was 0.25 (dashed line), but one sequential sample (squares) approached but never reached this, while the other (circles) yielded an accurate estimate at n = 40

taxon occurs only rarely, even though it might be very abundant when it does occur. Artifacts that tend to occur in hoards, for example, have low ubiquity but could be very abundant when they occur at all. Despite evidence that some authors have presented to show strong correlations between NISP and ubiquity in particular instances, there is no *necessary* relationship between abundance and ubiquity and no theoretical reason to think that there should be.

7.5.3 Diversity

Diversity is another measure that has nothing to do with abundance except insofar as the abundances of different taxa are similar or different. A diverse population is one that is distributed fairly evenly among a fairly large number of classes or taxa. Populations that are dominated by one or two taxa are not diverse. Ecologists developed diversity measures as a way to summarize the structure of biological communities and niche breadth (Hurlbert 1971; Peet 1974), but application of these concepts to archaeology is not simple (Cowgill 1989; Cruz-Uribe 1988; Sullivan and Tolonen 1998).

Diversity nonetheless has a number of useful archaeological applications. It can be a good way to analyze specialization, for example. Sites or activity areas with low artifact diversity are more likely to have been specialized in particular tasks. Low diversity in faunal or plant assemblages could indicate very focused hunting, gathering, or farming strategies (e.g., Bonzani 1997), as contrasted with the high diversity associated with the "Broad Spectrum Revolution" that some archaeologists have seen as preceding the earliest domestications of plants and food animals (Flannery 1968; Stiner 2001).

The simplest measure of diversity, richness, is just the number of taxa represented in a sample, that is, how many classes in a nominal scale have non-zero counts. However, this is extremely sensitive to sample size. A small sample is unlikely to represent all the taxa in a population, especially rarer ones (Meltzer et al. 1992), and richness has no unbiased population estimator.

Consequently researchers interested in diversity have attempted to account for sample size or sampling fraction. One approach has been to conduct computer simulations of samples of size n drawn from hypothetical populations for comparison with the actual sample of size n (Kintigh 1984, 1989).

Another approach is to use a richness index that takes sample size into account:

$$dl = \frac{s-1}{\ln(n)}$$

where dl is the richness index, s is the number of taxa represented in the sample, and $\ln(n)$ is the natural logarithm (base e) of the number of individuals in the sample (Kruz-Uribe [1988] substitutes MNI for n, with the bias that entails, and uses the common logarithm [base 10] instead of the natural one). This richness index compensates somewhat for bias in small samples but, as noted above, estimating number of individuals is problematic, substituting MNI for n creates bias, and this measure unrealistically treats small samples as though they are just as reliable as large ones.

Other diversity measures attempt to account not only for the number of taxa represented, but also for how evenly they are represented. The assumption is that, for a given number of

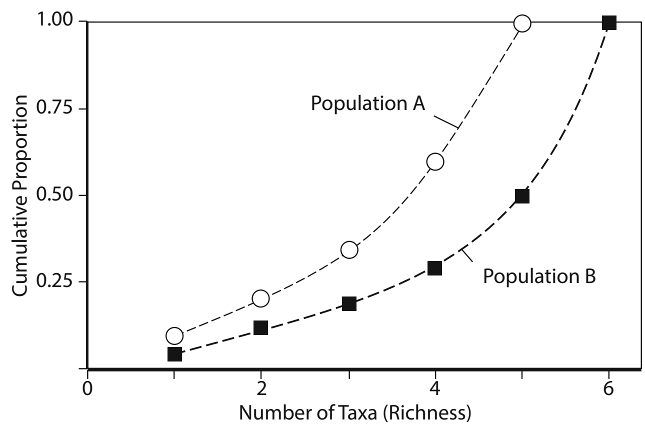


Fig. 7.7 When plotting richness (number of taxa) against the cumulative proportions of taxa, population *A* is more diverse than population *B* despite having fewer taxa because the cumulative proportions of taxa are higher (cf. Ringrose 1993b)

taxa, populations in which all or almost all taxa are fairly equally represented are more diverse than ones that display large variations in taxonomic abundance.

One way to address this is to compare populations' diversities graphically. If we rank taxa in increasing order of abundance, and then plot the number of taxa (s) on the x-axis against the cumulative proportion of individuals on the y-axis, we can compare a populations's richness and evenness simultaneously (Ringrose 1993b). In Fig. 7.7, for example, population A is more diverse than population B, despite having only five taxa instead of six, because 70% of the individuals belong to the two most abundant taxa in B, and the taxa in A, overall, are more evenly represented. However, this method is biased if we apply it to samples, rather than populations.

A common diversity measure is the Shannon-Weaver information function, or **Shannon index**:

$$D_{Sh} = \sum_{i=1}^{s} p_i \log_{10} p_i$$

This is just the sum over all s taxa of each taxon's proportion multiplied by the common logarithm of its proportion. We can estimate the proportions with measures like EVE or TMAU. For example, with three taxa with proportions of 0.5, 0.3 and 0.2, the Shannon index would be:

$$\begin{aligned} D_{Sh} &= 0.5*(-.3) + 0.3*(-.52) + 0.2*(-.7) \\ &= -.1 - .16 - .14 = 0.4 \end{aligned}$$

However, the Shannon index is biased because it is heavily influenced by the most numerous taxon. Hill (1973) proposes a way to decrease this dependence on most abundant taxa, but his approach also fails to provide an unbiased estimator (Kempton 1979; Ringrose 1993b).

Better indices are based on Simpson's index of dominance. When rearranged to express diversity instead of dominance, it is,

$$D_{Si} = 1 - \sum_{i=1}^{s} p_i^2$$

which is just 1 minus the sum of the squared proportions. Any population that has many taxa with similar abundances will have a lower value of the sum and therefore higher D_{Si} than would one dominated by one or two taxa. As you would expect, however, this index is also very sensitive to the abundance of the top one or two taxa, but it has the advantage of having an unbiased estimator for samples:

$$\widehat{D_{Si}} = 1 - \sum_{i=1}^{s} \frac{n_i(n_i - 1)}{n(n-1)}$$

Patil and Taillie (1979) propose yet another notation that unifies the richness, Shannon and Simpson indices yet has an unbiased estimator. We can express this family of indices as,

$$\Delta_\beta = \frac{1}{\beta}\left(1 - \sum_{i=1}^{s} p_i^{\beta+1}\right)$$

When $\beta = -1$, the equation describes a version of richness, s – 1, and when $\beta = 1$, the value of the Simpson index. Its unbiased estimator when β is a postive integer is,

$$\widehat{\Delta_\beta} = \frac{1}{\beta}\left(1 - \frac{\sum_{i=1}^{s} n_i(n_i-1)\ldots(n_i-b)}{n(n-1)\ldots(n-\beta)}\right)$$

Like the Shannon and Simpson indices, however, this is very sensitive to the proportions of the most numerous taxa (Ringrose 1993b).

Because the main problem with the most popular diversity measures is their lack of unbiased estimators for samples, Kintigh (1984, 1989) advocates sampling simulated populations to obtain a mean and variance for the diversity at a given sample size. However, it would be preferable to use an index as long as it provides unbiased estimators and is not too influenced by the most abundant taxa (Rhode 1988; Ringrose 1993b).

Smith et al. (1979) propose an "expected species" index that does exactly this. It is the expected number of taxa in a random sample of n individuals in a population:

$$s(m) = \sum_{i=1}^{s} \left(1 - (1 - p_i)^m\right)$$

for various values of m. When $m = 2$, the expected species is very similar to the Simpson index, and the most abundant one or two taxa have the greatest influence. Where $m > 2$, taxa of medium and minor abundance contribute more to the resulting value, and even more if m is much greater than 2. That is because $(1 - p_i)$, which is raised to the power of m, is largest for the taxa represented by small proportions. For samples, an unbiased estimator for $s(m)$ and integer values of m from 1 to n is,

$$\widehat{s(m)} = \sum_{i=1}^{s} \left(\frac{1 - C(n - n_i, m)}{C(n, m)}\right)$$

where $C(n, m) = \frac{n!}{(n-m)!m!}$ or the number of combinations of n items taken m at a time, and $n! = n(n-1)(n-2)\ldots1$. This is tedious to calculate manually for large samples, but spreadsheet software can handle it for relatively small samples and statistical or mathematical software for larger ones.

Ringrose (1993b: 282) suggests using this estimator of $s(m)$ to compare the diversities of archaeological samples by plotting a graph of m against $\widehat{s(m)}$ for all integer values between 2 and the smallest sample size (Fig. 7.8). The sample with the highest curve (in this case, A) is the more diverse (Birks and Line 1992).

As a final note, diversity in archaeological sample assemblages can vary because post-depositional differences among sites may result in greater destruction of certain taxa in some than in others. As with any of the measures discussed here, it is important to be sensitive to differences in the probability, r, that materials will survive to be included in the sample.

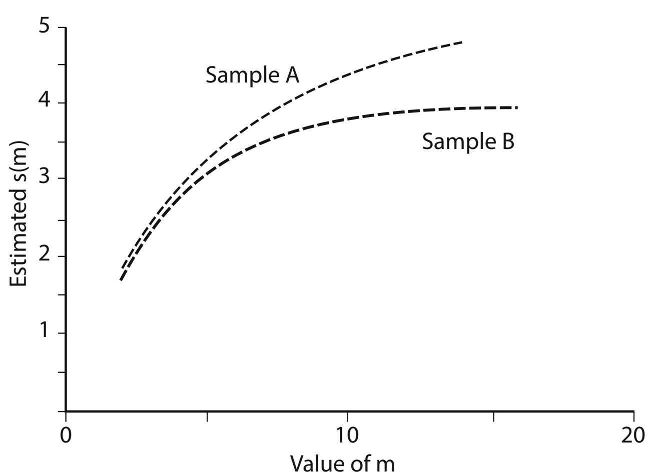


Fig. 7.8 Plot of m against the expected species index, $s(m)$. Sample A shows more diversity than sample B because its curve is higher (cf. Ringrose 1993b)

7.6 Summary

- Counts of fragments, bones, seeds and some other kinds of archaeological evidence do not have simple relationships to the numbers of whole pots, animals, plants or other entities
- Generally, these counts would be of little interest except as indirect measures of something that existed in the past
- Most abundance measures represent attempts to "correct" for some of the factors, such as numbers of parts in whole entities, degree of fragmentation, and probability of survival, that confound attempts at quantifying those past phenomena
- Some measures provide reasonable descriptions of samples, but are poor at characterizing past populations, except possibly at an ordinal scale
- Other measures purport to estimate population parameters, and have confidence limits, but generally present other challenges so that these confidence limits may be too optimistic
- Some aim to estimate or count numbers or amounts of things (e.g., Actual Number of Individuals, ANI), while others work well to estimate proportions of taxa
- Still others do not measure abundance at all, but instead measure a "scaling factor" (fractals), or whether distributions are concentrated or spread out evenly among spatial units (ubiquity) or among taxa (diversity)
- The best measures have unbiased estimators and confidence intervals, but even these require a skeptical eye

References Cited

Barrett, J. H. (1993). Bone weight, meat yield estimates and cod (Gadus morhua): A preliminary study of the weight method. *International Journal of Osteoarchaeology, 3*, 1–18.

Baxter, M. J. (2003). *Statistics in archaeology*. London: Arnold.

Baxter, M. J., & Cool, H. E. M. (1990). An approach to quantifying window glass. In K. Lockyear & S. Rahtz (Eds.), *CAA 90: Computer applications and quantitative methods in archaeology 1990* (BAR International Series 565) (pp. 127–131). Oxford: Tempus Reparatum.

Binford, L. R. (1978). *Nunamiut ethnoarchaeology*. New York: Academic.

Birks, H. J. B., & Line, J. M. (1992). The use of rarefaction analysis for estimating palynological richness from Quaternary pollen-analytical data. *Holocene, 2*(1), 1–10.

Bonzani, R. M. (1997). Plant diversity in the archaeological record: A means toward defining hunter-gatherer mobility strategies. *Journal of Archaeological Science, 24*, 1129–1139.

Boyd, W. E. (1988). Methodological problems in the analysis of fossil non-artifactual wood assemblages from archaeological sites. *Journal of Archaeological Science, 15*, 603–619.

Brown, C. T., Witschey, W. R. T., & Liebovitch, L. S. (2005). The broken past: Fractals in archaeology. *Journal of Archaeological Method and Theory, 12*(1), 37–78.

Byrd, J. E., & Owens, D. D. (1997). A method for measuring relative abundance of fragmented archaeological ceramics. *Journal of Field Archaeology, 24*, 315–320.

Casteel, R. W. (1974). A method for estimation of live weight of fish from the size of skeletal elements. *American Antiquity, 39*, 94–98.

Casteel, R. W. (1978). Faunal assemblages and the "weigenmethode" or weight method. *Journal of Field Archaeology, 5*, 71–77.

Cavanagh, W. G. (2009). Settlement structure in Laconia and Attica at the end of the Archaic period: The fractal dimension. *American Journal of Archaeology, 113*(3), 405–421.

Cavanagh, W. G., & Laxton, R. R. (1994). The fractal dimension, rank-size, and the interpretation of archaeological survey data. In I. Johnson (Ed.), *Methods in the mountains: Proceedings of UISPP commission IV meeting, Mount Victoria, Australia, August 1993* (Sydney University Archaeological Methods Series 2) (pp. 61–64). Sydney: Sydney University.

Chaplin, R. E. (1971). *The study of animal bones from archaeological sites*. London: Seminar Press.

Chase, P. G. (1985). Whole vessels and sherds: An experimental investigation of their quantitative relationship. *Journal of Field Archaeology, 12*(2), 213–218.

Chase, P. G., & Hagaman, R. M. (1987). Minimum number of individuals and its alternatives: A probability theory perspective. *OSSA, 13*, 75–86.

Corredor, D. M., & Vidal, J. M. (2016). Archaeological quantification of pottery: The rims count adjusted using the modulus of rupture. *Archaeometry, 58*(2), 333–346.

Cowgill, G. L. (1989). The concept of diversity in archaeological theory. In R. Leonard & G. Jones (Eds.), *Quantifying diversity in archaeology* (pp. 131–141). Cambridge: Cambridge University Press.

Cruz-Uribe, K. (1988). The use and meaning of species diversity and richness in archeological faunas. *Journal of Archaeological Science, 15*, 179–196.

Davis, M. B. (1963). On the theory of pollen analysis. *American Journal of Science, 261*, 897–912.

Davis, M. B. (1965). A method for determination of absolute pollen frequency. In B. G. Kummel & D. M. Raup (Eds.), *Handbook of palaeontological techniques* (pp. 647–686). San Francisco: Freeman.

Diehl, M. W. (2017). Paleoethnobotanical sampling adequacy and ubiquity: An example from the American Southwest. *Advances in Archaeological Practice, 5*(2), 196–203.

Drennan, R. D. (2010). *Statistics for archaeologists: A commonsense approach*. New York: Plenum Press.

Egloff, B. J. (1973). A method for counting ceramic rim sherds. *American Antiquity, 38*, 351–353.

Faegri, K., & Iversen, J. (1989). *Textbook of pollen analysis* (4h ed.). Oxford: Blackwell Scientific Publications.

Felgate, M. W., Bickler, S. H., & Murrell, P. R. (2013). Estimating parent population of pottery vessels from a sample of fragments: A case study from inter-tidal surface collections, Roviana Lagoon, Solomon Islands. *Journal of Archaeological Science, 40*, 1319–1328.

Fieller, N. R. J., & Turner, A. (1982). Number estimation in vertebrate samples. *Journal of Archaeological Science, 9*, 49–62.

Flannery, K. V. (1968). Origins and ecological effects of early domestication in Iran and the Near East. In P. J. Ucko & G. W. Dimbleby (Eds.), *The domestication and exploitation of plants and animals* (pp. 73–100). Chicago: Aldine Publishing.

Gautier, A. (1984). How do we count these bones, let me count the ways? Problems of archaeozoological quantification. In C. Grigson & J. Clutton-Brock (Eds.), *Animals and archaeology* (pp. 237–251). Oxford: British Archaeological Reports.

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Review of the Cambridge Philosophical Society, 41*, 587–640.

Grayson, D. K. (1978). Minimum numbers and sample size in vertebrate faunal analysis. *American Antiquity, 43*, 53–65.

Grayson, D. K. (1981). The effects of sample size on some derived measures in vertebrate faunal analysis. *Journal of Archaeological Science, 8*, 77–88.

Grayson, D. K. (1984). *Quantitative zooarchaeology*. Orlando: Academic.

Hastorf, C. A. (1988). The use of paleoethnobotanical data in prehistoric studies of crop production, processing, and consumption. In C. A. Hastorf & V. S. Popper (Eds.), *Current paleoethnobotany: Analytical methods and cultural interpretations of archaeological plant remains* (pp. 119–144). Chicago: University of Chicago Press.

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology, 54*, 427–432.

Hodder, I., & Orton, C. (1976). *Spatial analysis in archaeology*. Cambridge: Cambridge University Press.

Holtzman, R. C. (1979). Maximum likelihood estimation of fossil assemblage composition. *Paleobiology, 5*, 77–89.

Howard, H. (1930). A census of the Pleistocene birds of Rancho La Brea from the collections of the Los Angeles Museum. *Condor, 32*, 81–88.

Hubbard, R. N. L. B., & Clapham, A. (1992). Quantifying macroscopic plant remains. *Review of Palaeobotany and Palynology, 73*(1–4), 117–132.

Hulthèn, B. (1974). On choice of element for determination of quantity of pottery. *Norwegian Archaeological Review, 7*, 1–5.

Hurlbert, S. H. (1971). The nonconcept of species diversity: A critique and alternative parameters. *Ecology, 52*, 577–586.

Jackson, H. E. (1989). The trouble with transformations: Effects of sample size and sample composition on meat weight estimates bsed on skeletal mass allometry. *Journal of Archeological Science, 16*, 601–610.

Kadane, J. B. (1988). Possible statistical contributions to paleoethnobotany. In C. A. Hastorf & V. S. Popper (Eds.), *Current paleoethnobotany: Analytical methods and cultural interpretations of archaeological plant remains* (pp. 206–214). Chicago: University of Chicago Press.

Kempton, R. A. (1979). The structure of species abundance and measurement of diversity. *Biometrics, 35*, 307–321.

Kintigh, K. (1984). Measuring archaeological diversity by comparison with simulated assemblages. *American Antiquity, 4*, 44–54.

Kintigh, K. (1989). Sample size, significance and measures of diversity. In R. Leonard & G. Jones (Eds.), *Quantifying diversity in archaeology* (pp. 25–36). Cambridge: Cambridge University Press.

Krantz, G. S. (1968). A new method of counting mammal bones. *American Journal of Archaeology, 72*, 286–288.

Kruz-Uribe, K. (1988). The use and meaning of species diversity and richness in archaeological faunas. *Journal of Archaeological Science, 15*, 179–196.

Laxton, R. R., & Cavanagh, W. G. (1995). The rank-size dimension and the history of site structure for survey data. *Journal of Quantitative Anthropology, 5*(4), 327–358.

Lie, R. W. (1980). Minimum numbers of individuals from osteological samples. *Norwegian Archaeological Review, 13*, 24–30.

Lyman, R. L. (1979). Available meat from faunal remains: A consideration of techniques. *American Antiquity, 44*, 536–546.

Lyman, R. L. (1984). Bone density and differential survivorship of fossil classes. *Journal of Anthropological Archaeology, 3*, 259–299.

Lyman, R. L. (2008). *Quantitative paleozoology*. Cambridge: Cambridge University Press.

Lyman, R. L. (2015). On the variable relationship between NISP and NTAXA in bird remains and in mammal remains. *Journal of Archaeological Science, 53*, 291–296.

Lyman, R. L. (2018). A critical review of four efforts to resurrect MNI in zooarchaeology. *Journal of Archaeological Method and Theory*. org.myaccess.library.utoronto.ca/; https://doi.org/10.1007/s10816-018-9365-3.

MacNeish, R. S. (1967). A summary of the subsistence. In D. Byers (Ed.), *Environment and subsistence: Prehistory of the Tehuacan valley* (Vol. 1, pp. 290–309). Austin: University of Texas.

Marshall, F., & Pilgram, T. (1993). NISP vs. MNI in quantification of body-part representation. *American Antiquity, 58*, 261–269.

Mayer-Oakes, W. J., & Portnoy, A. W. (1993). Paleo-Indian studies at San José, Ecuador. *Lithic Technology, 18*, 28–36.

Meltzer, D. J., Leonard, R. D., & Stratton, S. K. (1992). The relationship between sample size and diversity in archaeological assemblages. *Journal of Archaeological Science, 19*, 375–387.

Miller, N. F. (1988). Ratio in paleoethnobotanical analysis. In C. A. Hastorf & V. S. Popper (Eds.), *Current paleoethnobotany: Analytical methods and cultural interpretations of archaeological plant remains* (pp. 72–85). Chicago: University of Chicago Press.

Minnis, P. E. (1981). Seeds in archaeological sites: Sources and some interpretive problems. *American Antiquity, 46*, 143–152.

Moore, P. D., Webb, J., & Collinson, M. C. (1991). *Pollen analysis: A laboratory manual*. Oxford: Blackwell.

Moreno-Garcia, M., Orton, C., & Rackham, J. (1996). A new statistical tool for the comparison of animal bone assemblages. *Journal of Archaeological Science, 23*, 437–453.

Morlan, R. E. (1983). Counts and estimates of taxonomic abundance of faunal remains: Microtine rodents from Bluefish Cave 1. *Canadian Journal of Archaeology, 7*, 61–76.

Noddle, B. A. (1973). Determination of the body weight of cattle from bone measurements. In J. Matolsci (Ed.), *Domestikationsforschung und geschichte der haustiere* (pp. 377–389). Budapest: Adadémiai Kiadó.

Orton, C. (1980). *Mathematics in archaeology*. London: Collins.

Orton, C. (1993). 'How many pots make five?' — An historical review of pottery quantification. *Archaeometry, 35*, 169–184.

Orton, C. (2000). *Sampling in archaeology*. Cambridge: Cambridge University Press.

Orton, C., & Tyers, P. A. (1992). Counting broken objects: The statistics of ceramic assemblages. *Proceedings of the British Academy, 77*, 163–184.

Orton, C., & Tyers, P. A. (1993). *A user's guide to pie-slice*. London: Institution of Archaeology, University College London.

Orton, C., Tyers, P. A., & Vince, A. G. (1993). *Pottery in archaeology*. Cambridge: Cambridge University Press.

Patil, G. P., & Taillie, C. (1979). An overview of diversity. In J. F. Grassle, G. P. Patil, W. K. Smith, & C. Taillie (Eds.), *Ecological diversity in theory and practice* (pp. 3–27). Fairland: International Co-operative Pub. House.

Peet, R. K. (1974). The measurement of species diversity. *Annual Review of Ecology and Systematics, 5*, 285–307.

Plug, C., & Plug, I. (1990). MNI counts as estimates of species abundance. *The South African Archaeological Bulletin, 45*(151), 53–57.

Poplin, F. (1976). Rémarques théoretiques et pratiques sur les unites utilisées dans les études d'ostéologie quantitative, particulièrement en archéologie préhistorique. In *Union International des Sciences Préhistoriques et Protohistoriques, 9e Congrès, Thèmes Spécialisées* (pp. 124–141). Nice: UISPP.

Popper, V. S. (1988). Selecting quantitative measurements in paleoethnobotany. In C. A. Hastorf & V. S. Popper (Eds.), *Current paleoethnobotany: Analytical methods and cultural interpretations of archaeological plant remains* (pp. 53–71). Chicago: University of Chicago Press.

Prange, H. D., Anderson, J. F., & Rahn, H. (1979). Scaling of skeletal mass to body mass in birds and mammals. *American Naturalist, 113*, 103–122.

Rackham, D. J. (1986). Assessing the relative frequencies of species by the application of a stochastic model to a zooarchaeological database. *PACT, 14*, 185.

Reitz, E. J., Quitmyer, I. R., Hale, H. S., Scudder, S. J., & Wing, E. S. (1987). Application of allometry to zooarchaeology. *American Antiquity, 52*, 304–317.

Reynolds, W. W. (1977). Skeletal weight allometry in aquatic and terrestrial vertebrates. *Hydrobiologist, 56*, 35–37.

Rhode, D. (1988). Measurements of archaeological diversity and the sample-size effect. *American Antiquity, 53*, 708–716.

Ringrose, T. J. (1993a). Bone counts and statistics: A critique. *Journal of Archaeological Science, 20*, 121–157.

Ringrose, T. J. (1993b). Diversity indices and archaeology. In J. Andresen, T. Madsen, & I. Scollar (Eds.), *Computing the past. Computer applications and quantitative methods in archaeology* (CAA92) (pp. 279–285). Århus: Århus University Press.

Rodriguez, E. C., & Hastorf, C. A. (2015). Calculating ceramic vessel volume: An assessment of methods. *Antiquity, 87*(338), 1182–1190.

Shott, M. (2000). The quantification problem in stone-tool assemblages. *American Antiquity, 65*(4), 725–738.

Shotwell, A. J. (1955). An approach to the paleoecology of mammmals. *Ecology, 36*, 327–337.

Smith, B. D. (1975). Toward a more accurate estimation of the meat yield of animal species at archaeological sites. In A. T. Clason (Ed.), *Archaeozoological studies* (pp. 99–106). Amsterdam: North-Holland Publishing Company.

Smith, W. K., Grassle, J. F., & Kravitz, D. (1979). Measures of diversity with unbiased estimates. In J. F. Grassle, G. P. Patil, W. K. Smith, & C. Taillie (Eds.), *Ecological diversity in theory and practice* (pp. 177–191). Fairland, MD: International Co-operative Publishing.

Solheim, W. G. (1960). The use of sherd weights and counts in the handling of archaeological data. *Current Anthropology, 1*, 325–329.

Stahl, P. W. (1982). On small mammal remains in archaeological contexts. *American Antiquity, 47*, 822–829.

Stiner, M. (2001). Thirty years on the 'Broad Spectrum Revolution' and Paleolithic demography. *Proceedings of the National Academy of Science, 98*(13), 6993–6996.

Stock, C. (1929). A census of the Pleistocene mammals of Rancho La Brea, based on the collections of the Los Angeles Museum. *Journal of Mammalogy, 10*, 281–289.

Sullivan, A. P., & Tolonen, A. S. (1998). Evaluatng assemblage diversity measures with surface archaeological data. In A. P. Sullivan (Ed.), *Surface archaeology* (pp. 143–155). Albuquerque: University of New Mexico Press.

Turcotte, D. L. (1986). Fractals and fragmentation. *Journal of Geophysical Research, 91*(B2), 1921–1926.

Van Pool, T. I., & Leonard, R. D. (2011). *Quantitative analysis in archaeology*. New York: Wiley-Blackwell.

Wallis, S. A. (2013). Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods. *Journal of Quantitative Linguistics, 20*(3), 178–208.

Wild, C. J., & Nichol, R. K. (1983). Estimation of the original number of individuals from paired bone counts using estimators of the Krantz type. *Journal of Field Archaeology, 10*, 337–344.

Wolverton, S., Dombrosky, J., & Lyman, R. L. (2016). Practical significance: Ordinal scale data and effect size in zooarchaeology. *International Journal of Osteoarchaeology, 26*, 255–265.

8 Probability, Modelling, and Statistical Inference

Sometimes apparently interesting patterns can arise purely by chance . . . or as a by-product of particular techniques of excavation or recording

Orton (1982: 214).

There would be little point in measuring archaeological materials if we did not use those observations to draw interesting inferences about what happened in the past, or what people's lives were like, whether at a particular site or more generally. Making such inferences convincingly requires, not only careful attention to measurement and error, as outlined in previous chapters, but also making those data work for us to make meaningful comparisons either between sites, time periods, activity areas or the like, or against what we would expect to observe if some hypothesis were true. The latter type of inference involves what some would call "verification" or "hypothesis testing."

Most of the time, archaeologists can only really say that one hypothesis is more probable than another — certainty is pretty rare in archaeology — so it is not surprising that probability and inferential statistics have some importance in our field. Even archaeologists who do not explicitly make use of statistical tests — even ones openly hostile to statistics, for that matter — often construct arguments that implicitly are statistical: this type of artifact is more common at this kind of site than at another kind, or an increase in marine mammals in these faunal assemblages over time is consistent with some hypothesis about changes in hunting strategies or subsistence economy, for example. Whether or not we acknowledge it, probability pervades all kinds of archaeology.

This chapter will review probability and some basic ways to test or evaluate (statistical) statements. In no way is it a substitute for thorough statistical training, and I would encourage those who do not already have a statistics background to consult some of the existing literature on archaeological statistics (e.g., Baxter 1994; Buck et al. 1996; Drennan 2010; Fletcher and Lock 1991; Shennan 1997) or, for those who need a very gentle introduction, Rowntree's (2018) *Statistics Without Tears*. The main purpose of this chapter is not to train readers in the use of particular statistical tests, but rather to familiarize them with probabilistic ways of thinking about data and hypotheses, and especially models.

8.1 "Verifying" Hypotheses

A "classical" approach to deciding whether a hypothesis adequately explains the data before us is called the Hypothetico-Deductive (or H-D) method. This proceeds from a hypothesis to deduction of what consequences should obtain if the hypothesis is true, then to a test to see if those consequences exist or not.

For example, we might hypothesize that the Natufian complex of the Middle Eastern Epipalaeolithic was pre-agricultural. We might deduce from that hypothesis that it had no domesticated crop plants. This is a valid deduction if we define agricultural societies as ones that have any domesticated plants in their economies, whether or not they grew them themselves (possibly not the best definition, but it will serve for demonstration), and if we accept that there are no "domesticated" plants at all in wild stands (which is not really true). To test the hypothesis, we would then conduct excavations at several Natufian sites, carefully selecting ones with no overlying Neolithic or later deposits that might contaminate the site with evidence of domesticates and using sieving and flotation to recover any preserved plant remains. By the reasoning of the H-D method, if we find any evidence at all of morphologically domesticated crop plants among the charred plant remains in these Natufian deposits, such as wheat or barley with tough rachis (see pp. 277, 286),

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_8

this would falsify the hypothesis that the Natufian was pre-agricultural, as defined here. The H-D method is inductive, despite use of "deductive" in its name. That is because failure to reject the hypothesis does not prove that the hypothesis is true; it only means it could be true, but other hypotheses may have equal or better claim to being true. For example, failure to find domesticates in a Natufian site could be due to poor preservation, unlucky sampling of sites, or error in the recovery or identification of the plant remains. These are all competing hypotheses that we also need to test if we are to make a convincing argument that the Natufian was pre-agricultural.

In any case, most archaeological tests of hypotheses do not even fit the H-D model, as just described. That is because they do not involve any deduction of things that *must* obtain if the hypothesis is true, only things that *will probably* obtain (Salmon 1982: 39–41; see p. 86). This is the form of statistical statements, and a common statistical method called **null-hypothesis significance testing** (NHST) involves testing the "null hypothesis," or the hypothesis that there is no difference between two samples or no differences from random data.

In evaluating statistical hypotheses, we must balance the risk that we will incorrectly reject a true hypothesis with the risk that we will incorrectly accept a false one. Rejecting a true hypothesis is called a **type I** (or α) error, while failing to reject a false hypothesis is called a **type II** (or β) error.

8.2 Probability

Most of us have at least basic understanding of probability as we use it in our daily lives, if only to interpret weather reports or to play well at card games. In simplest terms, probabilities are proportions that represent, over the long run, how often we would expect some event to occur "by chance." For example, if we knew that, in some population, the proportion of pot sherds with red paint on their exterior is 0.30 (or 30%), then the probability that any particular sherd we pull at random from that population will have red paint is also 0.30. This is just like our knowledge that the probability of pulling a black card from a fair deck of 52 cards is 0.50, and the probability of drawing a Queen is 4/52 = 0.077. We also understand that when there are only two possible outcomes of equal probability, such as getting heads or tails on a coin flip (landing on the coin's edge having negligible probability), that probability is 0.5. Similarly, in an archaeological assemblage of animal bones, for those bones that have left and right elements, we expect that the probability that any given bone drawn at random is from the left side of the body is also 0.5.

This does not mean that we can know how any particular random draw will turn out. All we can say is that, if we repeatedly drew bones from a very large assemblage randomly, *about* 50% of them would be left bones. Furthermore, the longer we make such drawings, the closer that percentage would approach 50%. The proportion that our sample approaches with increasing sample size is called the **limit** — the proportion of "successful" trials in an extremely large sample.

In cases like coin flipping and left and right faunal elements, we have *a priori* knowledge of what the proportion of "heads" or "left" should be in a **population**. On theoretical grounds, we know in these cases that it should be 0.50, and that would be the limit in a very large sample of coin-flipping experiments. In many archeological cases, we can only guess the population proportion or estimate it from a sample (see "Sampling," pp. 94–102).

8.2.1 Properties of Probabilities

The following might seem a little daunting to some, but even readers without statistical background should try to follow it. The basic principles of probability are really common sense.

First, because they are proportions, probabilities always range from 0 to 1.0. If we represent the probability of an event called A as $p(\mathrm{A})$, then,

$$0 \leq p(\mathrm{A}) \leq 1$$

Whenever $p(\mathrm{A}) = 1$, event A is certain to occur. Conversely, $p(\mathrm{A}) = 0$ means that event A is impossible or will never occur.

Addition Rule Whenever two events, A and B, are mutually exclusive (e.g., a coin flip cannot be heads and tails simultaneously), then we can add those probabilities. For example, the probability that a roll of a single die will produce either a 1 or a 2 is 1/6 + 1/6 = 1/3. We can express this as,

$$p(\mathrm{A\ or\ B}) = p(\mathrm{A}) + p(\mathrm{B}).$$

The sum of probabilities of all mutually exclusive outcomes must be 1.0. Consequently, the probability that some event will *not* occur is the sum of all the other outcomes, or,

$$p(\mathrm{notA}) = 1 - p(\mathrm{A})$$

For example, if we knew that the probability that any projectile point from a population was notched was 0.4, then the probability of drawing an un-notched projectile point at random from that population must be 1–0.4 = 0.6.

However, the addition rule does not apply in cases where the outcomes could occur at the same time. For example, the probability that a given potsherd comes from a jar (event A) does exclude the possibility that it has red paint (event B). For cases like this, we must modify the addition rule as follows, to describe the case that a particular sherd will be from a jar or has red paint:

$$p(\text{A or B}) = p(\text{A}) + p(\text{B}) - p(\text{A and B})$$

The last term just prevents us from double-counting sherds that are both red-painted and jars. For cases with more than two kinds of outcome that are not mutually exclusive, this notation gets a bit more cumbersome. For three outcomes:

$$\begin{aligned} p(\text{A or B or C}) = p(\text{A}) + p(\text{B}) + p(\text{C}) - p(\text{A and B}) \\ - p(\text{A and C}) - p(\text{B and C}) \\ + p(\text{A and B and C}) \end{aligned}$$

Multiplication Rule When we have possible outcomes that are not mutually exclusive, we are likely to be interested in how often they will co-occur. For example, how likely is it that a jar will have red paint? The probability of co-occurrence is equal to the probability of one of the outcomes times the **conditional probability** of the other. The conditional probability is the probability that the second event will occur given that the first has occurred. We represent the "given that" expression as "|" so,

$$p(\text{A and B}) = p(\text{A}) * p(\text{B}|\text{A}) = p(\text{B}) * p(\text{A}|\text{B})$$

Note that there are two ways to arrive at p(A and B). Whether we start with p(A) or p(B), we get the same result. This might be clearer to you if you remind yourself that probabilities are essentially proportions. If 15% of the pottery in an assemblage comes from jars, and 5% of the pottery in that assemblage is red-painted, then $p(\text{A}) = 0.15$, $p(\text{B}) = 0.05$. Let's say that 20% of the jars have red paint, so $p(\text{B}|\text{A}) = 0.2$. Therefore, if we randomly selected a sherd from the assemblage, the probability that it is from a red-painted jar, p(A and B), would be $0.15 \times 0.2 = 0.03$. From this, we can also re-arrange terms to infer the probability that any red-painted sherd is from a jar:

$$p(\text{A}|\text{B}) = p(\text{A and B})/p(\text{B}) = 0.03/0.05 = 0.6$$

So, we can infer that about 60% of the red-painted sherds come from jars.

The relationship that the multiplication rule exposes brings up the notion of **independence**. We can consider two outcomes or events to be independent if the occurrence of one in no way affects the occurrence of the other. Mathematically, if A and B are independent, then $p(\text{A}|\text{B}) = p(\text{A})$ and $p(\text{B}|\text{A}) = p(\text{B})$. In other words, the condition after the "|" sign has no effect whatsoever. In the pottery example above, the probability that a sherd will be a jar is 0.15, but the conditional probability that it is a jar, given that it is red-painted, is 0.2. As these are not equal, it is obvious that pot shape and painting are not independent. This suggests the possibility that there is an interesting pattern in the distribution of pottery treatments, and is the basis for some very common statistical tests, such as the chi-square test.

In cases where we are sure that outcomes are independent, however, we can simplify the multiplication rule as follows:

$$p(\text{A and B}) = p(\text{A}) * p(\text{B})$$

For example, since we know on theoretical grounds that the color of a playing card is independent of its value, the probability of getting a red Queen, since half the deck is either Hearts or Diamonds, would be $½ \times 1/13 = 1/26 = 0.038$. Unlike card playing, however, most archaeological research does not allow us to make *a priori* assumptions of independence.

8.3 Models

A lot of statistical inference is based on comparing some outcome with the outcome we would expect to have if our sample had been drawn randomly from a particular kind of theoretical distribution. It does not necessarily mean that we expect any particular archaeological results to match one of these distributions; they are just useful standards for comparison. These distributions are **models**. Among the things we must consider when selecting a model is whether the data of interest are discrete or continuous, limited to non-negative values, or limited to a specific range of values.

We can model coin-flipping, for example, with the **binomial model**. This applies to cases where we have a fixed number of independent observations or trials, and each trial can have only one of two outcomes, such as heads or tails, and the number of any outcome, say heads, cannot exceed the number of trials (Buck et al. 1996: 91–92; Doran and Hodson 1975: 45–48). In the case of coin-flipping, we can assume *a priori* probabilities of 0.5 for each of the possible outcomes and we would expect, in the long run of thousands of coin flips, to get roughly equal numbers of heads and tails.

The parameters of a binomial distribution are n (the number of trials, such as coin flips) and p (the probability of getting a particular outcome on each trial). For ten coin flips, $n = 10$ and $p(\text{Heads}) = 0.5$. We can summarize this particular binomial model as B(10,0.5), meaning, "Binomial distribution with ten trials and a probability of 0.5 of getting Heads in any one trial," and the mean or expected value of a binomial distribution is np, in this case, $10 \times 0.5 = 5$. The standard deviation of a binomial distribution is:

$$\sigma = \sqrt{np(1-p)}$$

Of course, the value of p does not have to be 0.5. A dishonest coin, constructed to favor Heads, for example, might have $p(\text{Heads}) = 0.6$ and $p(\text{Tails}) = 0.4$. While the distribution for a probability of 0.5 is symmetrical, any other probability yields an asymmetrical distribution (Fig. 8.1).

There are many archaeological cases that we could reasonably model with a binomial distribution. Any measure of ubiquity, as in paleoethnobotanical applications (see Chap. 7), involves a fixed, discrete number of observations that either do or do not include a particular taxon. For example, perhaps the *a priori* probability of finding at least one maize kernel in each 1 m^2 excavation unit at a site is 0.4. Let us suppose that we excavate ten of these 1 m^2 units. In that case, we have B(10,0.4) but it is conceivable that our excavation might only find maize in one of these units, that is, ubiquity is $1/10 = 0.1$. Given the *a priori* probability of 0.4, we might have expected to find maize in $10 \times 0.4 = 4$ units. Is the deviation from our expectation large enough to suggest that something unexpected is going on? Perhaps maize preservation is unusually poor or we excavated in the wrong place? Or is the result just a function of the probabilistic nature of our sample?

Examination of the binomial distribution for this case (Fig. 8.1) would indeed suggest that it is very unlikely that we would find maize in only one unit if the a priori probability of 0.4 is correct. In fact, the probability of finding maize in only one of ten units if the probability is 0.4 is only 0.0404, or less than 5%. For B(10,0.4), the standard deviation (for a sample) is:

$$s = \sqrt{10 * 0.4 * 0.6} = 1.55$$

What this means is that we would have expected the number of units with maize to have been 4 ± 1.55. Below, we will consider what happens when we find a case that does not meet this expectation.

Another useful model is the **Poisson distribution**, sometimes called the model for rare events. This models discrete (countable) random events per unit of time or space and, unlike the case for the binomial model, there is no upper limit on the number of these events, but the lower end cannot have values less than 0 (Buck et al. 1996: 92–95; Doran and Hodson 1975: 48–49). It is the distribution that models counts of ^{14}C disintegrations per minute in the older, decay-counting method of radiocarbon dating. It also models the number of potsherds present per square meter on or off a site, or obsidian flakes per cubic meter in an excavation.

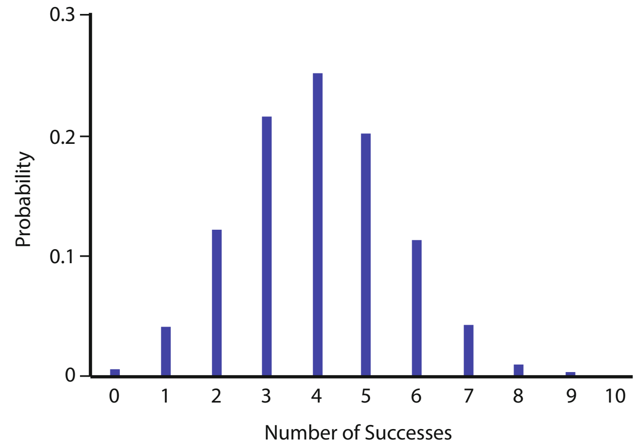


Fig. 8.1 Binomial model for a probability of 0.4 of finding at least one maize kernel in ten 1 m^2 excavation units, B(10,0.4). Note that the most likely outcome is that four units will have one or more maize kernels, and there is very little chance that none, or as many as eight, will have a maize kernel. The bars are separate line segments to signal that this is a discrete distribution

The Poisson model has only one parameter, λ, which is both the mean (or density) and the variance. Consequently, the standard deviation is $\sqrt{\lambda}$. The probability of finding x things, where $e = 2.718\ldots$, and $x!$ (x factorial) is the product of all integers from 1 to x, is:

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2, 3, \ldots$$

For example, in a sample of 1 m^2 units, and a density of 3 artifacts per square meter, the probability of finding two artifacts in one unit is,

$$p(2) = \frac{e^{-3}3^2}{2 * 1} = \frac{2.718^{-3} * 3^2}{2 * 1} = \frac{0.448}{2} = 0.224$$

Doing this calculation for integer values of x from 0 onward results in a graph like that in Fig. 8.2, which shows the distribution for $\lambda = 5$. Typically, except when λ is very large, Poisson distributions are markedly asymmetrical. When λ is extremely large, as when accelerator mass spectrometry is used to count thousands or millions of carbon atoms for radiocarbon dating, the shape of their distribution (number of atoms per minute) becomes more symmetrical and approaches the shape of a normal distribution.

As Kadane (1988) points out, recognizing that many archaeological observations may fit the Poisson model has important implications for such things as optimizing sample sizes and interpreting the effects of differential preservation. Furthermore, even when the process in which archaeologists are interested is probably not Poisson, sometimes it makes sense to compare it with a Poisson one. For example, we often expect archaeological artifacts to be clustered in space

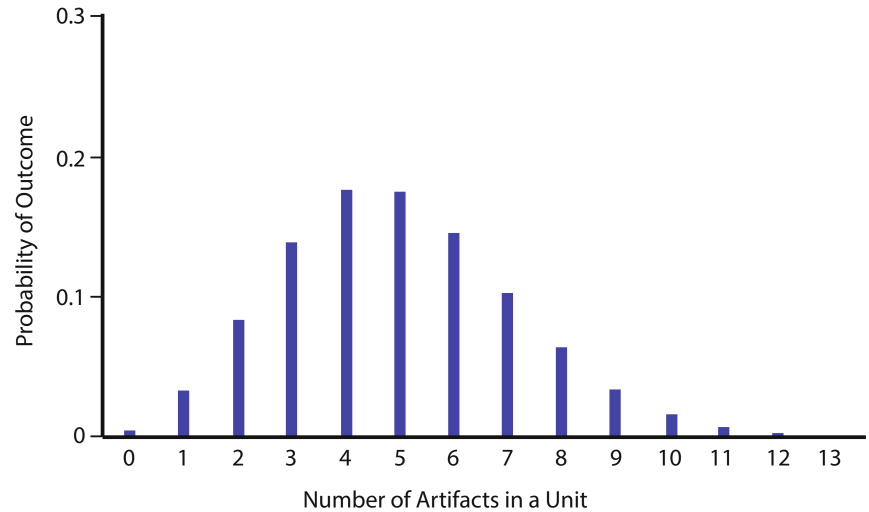


Fig. 8.2 Poisson distribution for a density of 5 artifacts per square meter ($\lambda = 5$)

but, if that is so, we should expect artifact distributions (as represented as artifacts per unit space) to deviate from the Poisson model, since the latter describes distributions that are random in space.

One very useful model is the **normal distribution**, also called the **Gaussian distribution**. This models the behavior of repeated measurements of some interval- or ratio-scale phenomenon, such as the lengths of blades or the mass of spindle whorls. It has no upper or lower limits on its values and is applicable to continuous data. This is the familiar "bell curve," centered on the mean or average (Fig. 8.3). The normal distribution becomes narrower as the sample size on which it is based becomes larger, meaning that its standard error decreases with sample size.

Interestingly, if we take repeated samples of size n from some population, even if that population is *not* normally distributed, the distribution of all the sample means from those samples will always approach a normal distribution with a mean of μ and variance of σ^2/n. This useful property is called the **central limit theorem**, and it means that we can use the normal distribution even to evaluate populations that are not themselves normally distributed. It happens because you would expect samples from a population to have values that mostly cluster near the mean — obtaining very many extreme values, especially extremes consistently in the upper or lower part of the distribution, would be very unlikely — and also because the means of these samples' means would cluster even more strongly near the population mean. This also results in less dispersion than in the original population.

Another useful property is that this normal distribution of sample means is a probability distribution. That means that the area under any portion of the curve represents the sum of the probabilities of all the outcomes in that range of the x-axis on the graph (Fig. 8.3). If, for example, we make repeated samples of projectile point lengths and graph their sampling distribution, then draw a vertical line from the 5 cm mark on

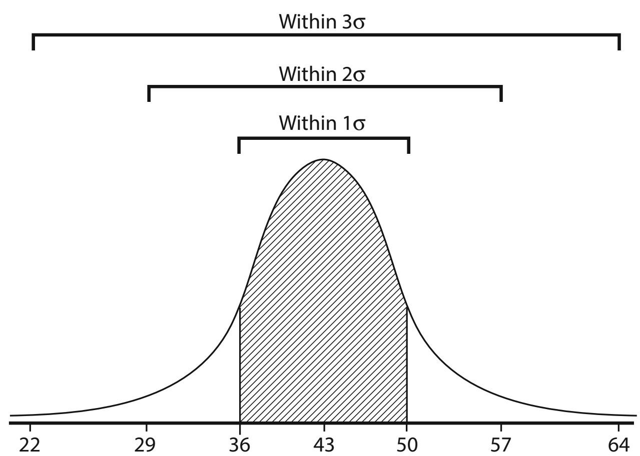


Fig. 8.3 Normal or Gaussian distribution for the lengths of a population of projectile points with a mean length of 43 mm and standard deviation (σ) of 7 mm. The hatched area represents slightly more than 68% of the area under the curve, and each tail represents almost 16%. The area within 2σ is about 95%, and within 3σ is about 99.7% of the curve's total area

the x-axis, the area under the curve to the left of this line represents the probability that any projectile point you randomly draw from the population will be less than 5 cm long, and the area to its right is equivalent to the probability that it will be longer than 5 cm. By drawing two such lines instead of one, we can determine, by area, the probability of any range of outcomes. For example, the corresponding area might show that there is a probability of 0.4 that any random projectile point will be more than 6 but less than 10 cm long.

There are other useful models, some of which you will encounter in this text, such as **student's t**, **beta**, **gamma**, **log-normal**, and **hypergeometric distributions** (Buck et al. 1996: 83–113). Some of these are very similar to ones just described, but differ in that they are for continuous, rather than discrete, data or have larger standard deviations. The gamma distribution, for example, is a continuous version of the Poisson distribution; rather than model discrete, countable events in space or time, we could use it to model things like the average *distance* in space or time between discoveries of artifacts along a survey transect. Some other distributions, such as the exponential and chi-square distributions, are based on the gamma distribution. Student's t-distribution is a very useful one as it is very similar to the normal model but has a larger standard error when sample size is small. It is thus the one we should be using for relatively small sample sizes; for sample sizes much greater than about 50, the t-distribution is almost identical to the normal distribution.

8.4 Parametric Statistical Tests

Parametric tests are statistical tests that involve comparing measurements and evaluating hypotheses by reference to one of these theoretical models. In their classical version, this is null-hypothesis significance testing (NHST), which compares actual measurements to the expectations for a **null hypothesis** of no difference between samples or between a sample and a theoretical model. **Non-parametric tests**, such as the Mann-Whitney test and Spearman's rank correlation (Drennan 2010: 224–229), by contrast, are ones that do not require the assumption that the data fit one of these models (distribution-free) or that have a distribution but its parameters are not specified.

One of the parametric tests, the **Z-test**, takes advantage of the properties of the normal model of a sampling distribution to help us draw inferences about some sample or samples. For example, hypothetically, at a very large site, we might have a population of pit-houses whose mean length was known to be $\mu = 7.5$ m with a standard deviation of $\sigma = 1.5$ m. If we drew a random sample of 100 houses from this population and found the mean of this sample to be 7.9 m, is the difference due to some problem with our sample? Are our measurements biased? The characteristics of the normal distribution allow us to answer these questions as long as we can reasonably make some assumptions. We would assume that our sample was a random one (although really this is the hypothesis we are trying to assess). If that sample is not very large, we would also have to assume that the population was randomly distributed (we can relax this assumption here because a sample size of $n = 100$ is fairly large). We assume a continuous scale of measurement, which is fine for lengths measured in meters. Finally, since here we have privileged information, we assume that $\mu = 7.5$ m and $\sigma = 1.5$ m.

Next, we would employ the normal distribution as our sampling distribution. At one time, we would use tables in a book that tell us the probability values for various increments along the x-axis of the distribution; today we have computer software that can measure the areas under the normal curve for any range we specify. The units along the x-axis are "Z-scores," which are units of standard error. For example, in Fig. 8.4b, the shaded area in the right tail of the distribution is to the right of Z = 1.65 (or 1.65 standard errors from the mean), and corresponds to 0.05 of the entire area under the curve of the normal distribution.

Our working hypothesis is that there is something suspicious about our sampling methods, but the hypothesis we will

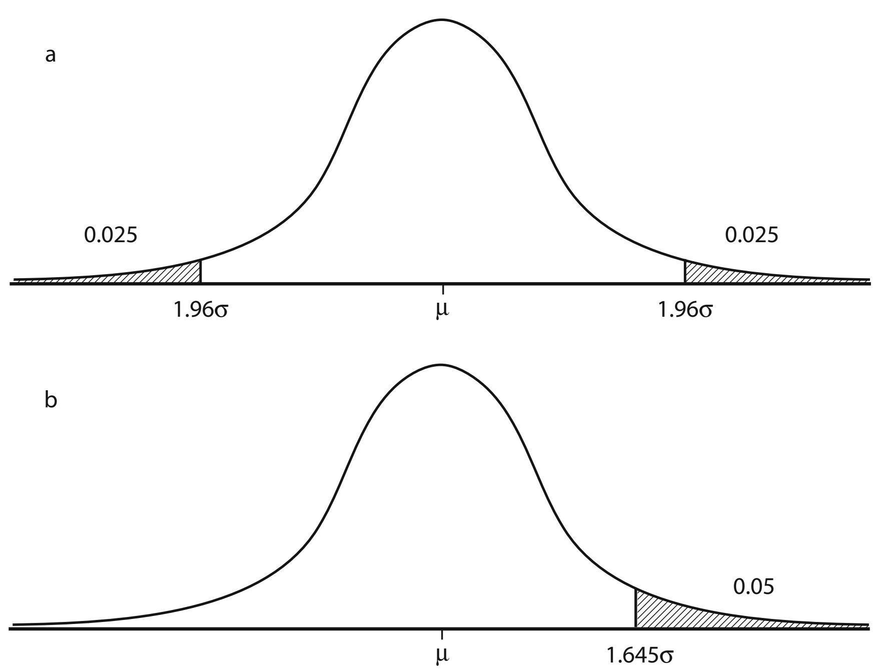


Fig. 8.4 Examples of a two-tail Z-test with two rejection regions that jointly add up to 0.05 of the curve's area (**a**) and a one-tail test, also with a significance of 0.05, but with all the area in one tail (**b**). In classical statistics, a result within one of the tails (hatched) would result in rejection of the null hypothesis that the mean = μ at 95% confidence

actually test is called the **null hypothesis** (sometimes represented as H_o) that there is no significant difference between the mean of our sample and a truly random sample from the same population. Mathematically, that means that there is no significant difference between the population mean, μ, and the sample mean, $\bar{x}$. Some alternative to the null might be our favored or just alternative hypothesis, sometimes represented as H_1 or H_A.

Next, we must decide what we mean by a "significant difference." This is the significance level, which involves our willingness to tolerate a **type I** or **type II error**. As mentioned above, a type I error is rejecting a true hypothesis, while type II involves failing to reject a false one. How we balance these potential errors calls on us to be "conservative." What that means is that we should select a significance interval that does not make it too hard for us to reject our favorite hypothesis (or that does not make it too easy to reject the null hypothesis, if we do not favor the null). If, for example, we pick a significance level of 0.05, that would mean we are willing to risk a type I error one time in twenty (or 95% confident that rejecting the hypothesis is justified). If, instead, we picked 0.1, we would risk a type I error one time in ten, but with less risk of a type II error.

Another decision we need to make is whether the test should be one-tailed or two-tailed. We would use a **one-tailed test** whenever we can predict the direction of difference from the null hypothesis. In the example of the pit-houses, for example, our alternative hypothesis is that our sample is giving us a biased estimate (we suspect we have oversampled large houses), so a one-tailed test is appropriate. That means, at 0.05 significance, that we would see if the result of our Z-test comes out in the upper 5% of the distribution (the rejection area). Had we no prediction as to whether our sample of house pits was too small or too large, we could use a two-tailed test, with rejection areas of 0.025 at both the left and right ends of the distribution.

We're then ready to do the test. Its mechanics are very simple. Essentially, we just convert whatever units we have for our interval/ratio measurements (in this case meters) to "Z-scores," which are units of standard error. Consequently, we first need to calculate the standard error of our sample of 100 pit-houses:

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.5}{\sqrt{100}} = \frac{1.5}{10} = 0.15$$

Next, we measure the difference between the mean house length in the sample and that in the population in terms of Z-scores:

$$Z = \frac{\bar{x} - \mu}{SE} = \frac{7.9 - 7.5}{0.15} = 2.67$$

For a one-tailed test, the rejection region for 0.05 significance would be to the right of $Z = 1.65$. 2.67 clearly exceeds this, in fact, by a wide margin, and even exceeds the Z-score for 0.01 significance: 2.33. Consequently, even at 0.01 significance, we can conclude that the house-pits in the sample are unusually big. This result might suggest that there was something wrong with our sampling method (perhaps we unwittingly used a version of PPS sampling, see pp. 97–98), and we would then try to find a better one or make appropriate adjustments.

This example illustrates what we might call the **classical** approach to statistics, which includes defining carefully, in advance, both the null hypothesis and at least one alternative hypothesis, and deciding on what result would result in our either accepting or rejecting the null. Note that it is still inductive because the risks of type I and type II errors mean that we can never be absolutely certain that our conclusion is correct. Our result could be an outlier or that there could be some hypothesis we did not consider that explains the result better than our alternative hypothesis does.

Not all statisticians or researchers favor this classical approach. Some skip the selection of a rejection area and just report the significance outcome that results from the test. In the example just mentioned, they would just say that the chance of obtaining this result by chance alone is less than 0.0038 (the area of the curve to the right of 2.67), and leave it to readers to decide whether or not they think this warrants rejection of the null hypothesis. Note that the significance level does *not* tell us the probability that the hypothesis is true or false, only the probability of getting a difference as large as we did in sample data if the hypothesis is true.

One practice that may seem like this last approach is not a valid one, despite its popularity in some quarters. This is what you might call a "fishing expedition" and involves conducting a large number of statistical tests on a whole bunch of variables just to see if any of the tests turn up "significant" results, such as probabilities less than 0.05 or 0.01 of obtaining the result we did if the null hypothesis is true. Because these tests do not begin with credible alternative hypotheses or careful research questions, there is always a possibility that the statistical result is just by chance. Remember: the very definition of a 0.05 significance level is that there is one chance in twenty of getting a Z-value that large even if the null hypothesis is true (a type I error). So, if you conduct 100 tests of randomly or arbitrarily selected null hypotheses, we can expect about five of them to give results in the 0.05 rejection area and one will likely fall in the 0.01 rejection area even if there is no pattern in the data at all. Always have a plausible alternative hypothesis in mind *before* you conduct a statistical test.

Yet another approach to statistical hypothesis evaluation is the **Bayesian** one. This approach does not just accept or reject hypotheses, it determines whether a particular hypothesis becomes weaker or stronger in the light of new evidence. It also allows us to compare the relative merits of competing hypotheses by reference to the way that new evidence affects them. We will return to Bayesian inference below (p. 139).

Returning to the classical approach, in this example of a **one-sample test** (comparison of a sample statistic to a known parameter), we were in the unusual position of knowing in advance what the population mean and standard deviation were. In real research, we are unlikely to know the population parameters (that is often what we are trying to estimate), and it is much more common for archaeologists to compare two samples to infer whether they may have come from the same population. This is called a **two-sample test**.

A common use for a two-sample test is to compare two means: the difference-of-means test. For example, we might want to know whether there are significant differences between the size or other attributes of projectile points from two sites. If there is not, they could be drawn from the same population (perhaps products of the same "community of practice" of knappers). For samples, we conventionally indicate the mean as $\bar{x}$ instead of μ, and the standard deviation as s instead of σ. If we use length as one measure of interest, our null hypothesis is that the means of samples drawn from the two sites are equal. The alternative hypothesis might just be that they are not equal (and thus call for a two-tail test) or, if we have already formed an opinion on what kind of difference there is, that the projectile lengths at site A are greater than those at site B, on average (one-tail test).

Example

Almost always, faunal remains from sites come from cluster samples (see pp. 98–101) of some type, and analysts often only report the proportion or number of each taxon in a whole assemblage (whole site or layer within site) in publications. In such cases, we do not have the raw data on how the distributions of bones varied across contexts at the sites. Let us imagine that one such report says that in a sample of 150 identifiable specimens at site A, 85 (or 57%) were red deer (*Cervus elaphus*), while at site B, 57 (or 46%) identifiable specimens in a sample of 124 were red deer. Are we justified in inferring that the inhabitants at site A hunted or consumed more deer? For the moment, we will ignore some complicating (or confounding) factors about the relationship of identifiable bone fragments to diet (but see Chap. 7), and just ask ourselves if a difference as large as the one we see here could just be due to chance. Using the classical NHST approach, the null hypothesis is that the two population proportions are the same and the alternative hypothesis might be that the proportion at site A is higher. Since we do not know the population proportions, we substitute what we do know, their estimates from samples, p_A and p_B. Since, in this case, we do not have enough information to calculate standard errors for cluster samples, we have to make do with estimating each standard error as follows:

$$\widehat{SE} = \sqrt{\frac{p * (1-p)}{m}}$$

where p is the proportion at one of the sites and m is the total number of identifiable bones and bone fragments from that site. For the reported proportions of 0.57 at site A and 0.46 at site B, we would thus estimate the two individual standard errors as 0.04 and 0.045.

To conduct the test, we are interested in the **distribution of differences** between the two samples. Since the variance of this distribution is equal to the sum of the two individual variances, we can just add them together and then take the square root, as we did for the difference-of-means, since variance is just the square of standard deviation:

$$SE_{A-B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$
$$= \sqrt{\frac{0.57 * 0.43}{150} + \frac{0.46 * 0.54}{124}} = 0.0609$$

If we assume that the two samples are random, independent samples, and select a rejection area corresponding with the 0.05 significance level for a one-tail test (because the alternative hypothesis is that the proportion at site A is higher than at site B), we are looking for a Z-score greater than 1.65 to reject H_o in favor of H_A. The test then becomes:

$$Z = \frac{p_A - p_B}{SE_{A-B}} = \frac{0.57 - 0.46}{0.0609} = 1.8$$

Since 1.8 exceeds 1.65, we may reject the null hypothesis with 95% confidence, meaning that we tentatively accept the alternative, H_A, that site A has a higher proportion of deer than site B, but with a risk of incorrectly accepting H_A (type I error) that corresponds with the area to the right of 1.8 (~0.036).

The Z-test for this situation would be very similar to that for a one-sample test except that we replace μ in the equation with a second sample mean:

$$Z = \frac{\overline{x}_A - \overline{x}_B}{SE_{A-B}}$$

Note that here we need to calculate SE, not on one of the samples, but on the distribution of differences[1] between the samples:

$$SE_{A-B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

This just means that we divide an estimate of the variances of sample A and B by their sample sizes, sum these, and then take the square root. We can also use two-sample tests to compare proportions. A very common situation in archaeology is to decide whether proportions in samples from different sites, or different layers or spaces within the same site, are very different. For example, we might want to compare the deer bones and bone fragments from two sites as proportions of all the bones and fragments at those sites. This would call for a **difference-of-proportions tes**t. This is much the same as the difference-of-means test except that it uses a different way to account for standard error.

Parametric statistical tests based on other kinds of distribution work much the same way as the Z-test. This is particularly true of the t-test, which is exactly the same as the Z-test except that the t-distribution is wider and flatter than the normal distribution when sample sizes are fairly small, but virtually identical to the normal distribution when the sample exceeds 100. If you are not sure if your sample size is large enough to warrant a Z-test, but your data meet the other assumptions (interval/ratio scale, independent random samples, equal variances), just use a t-test.

Testing a hypothesis that involves a fixed sample size and the presence or absence of something in each sample element, as in ubiquity measures (Chap. 7), would call for use of the binomial model instead of the normal or t-distributions.

Other useful archaeological applications of the Poisson distribution are in estimating the effectiveness of shovel-test surveys (e.g., Krakker et al. 1983) or the amount of sediment that we need to "float" to get reasonable numbers of plant remains (Lee 2012).

8.5 Confidence Intervals

Many archaeologists think of **confidence intervals** as the range within which the "true" value lies, at a certain level of confidence. However, this is an oversimplification. Technically, a 95% confidence interval, for example, does not mean that there is a 95% probability that the interval contains the "true" value. The 95% actually refers to the reliability of our estimate of the interval. If we were to take very many samples — usually we just take one — from a population, the proportion of samples whose confidence interval included the true parameter would be close to 95% (Neyman 1937; Morey et al. 2016). That is technically not the same thing as saying there is a 95% probability that the confidence interval of a single sample contains the parameter.

Example

Returning to the example where the density of maize at a site was such that the probability of finding at least one maize kernel in a 1 m^2 excavation unit is 0.4, if we were to excavate ten such units, the model would be B(10,0.4), as noted above. The expectation would be that four of the ten units should have at least one maize kernel (10 × 0.4). And, as already calculated above, he standard deviation for this case is $\sqrt{10 * 04 * 0.6} = 1.55$.

If we do not find any maize at all in nine of the units, is this then a surprising result? We can answer this question by reference to the graph of the distribution for B(10,0.4) in Fig. 8.1. The probability of finding maize in only one of ten excavation units is only 0.0404. Furthermore, the probability of finding maize in less than two such units is the sum of the probabilities for 1 and 0 units, 0.0403 + 0.0061 = 0.046. Consequently, this is a very unlikely outcome if the ubiquity of maize here is actually 0.4.

The equation for finding the probability of any of the values in Fig. 8.1, as for any binomial distribution, is:

$$p(x; n) = \frac{n! p^x (1-p)^{n-x}}{x!(n-x)!}$$

Where $p(x;\ n)$ is the probability of obtaining x successes in n trials, $n!$ (or n factorial) comes from multiplying all the integers from 1 to n, and p is the probability of a success on any one trial. This allows us to determine the probability of any outcome and reject or provisionally accept a null hypothesis based on the binomial model. Thankfully, statistical software automates tests based on the binomial distribution.

[1] The following equation simplifies by the assumption that the two samples have equal variances. For unequal variances, you should substitute n – 1 for n in the denominator of the SE ratio.

Example

We can similarly make use of the Poisson distribution to test hypotheses. Let us imagine that we have excavated part of a site in units that are 1 m^2 in area and 0.1 m in depth, so that each has a volume of 0.1 m^3. The probability that any such unit will contain a particular number (*x*) of lithic flakes greater than 1 cm in their long axis is:

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{for x} = 0, 1, 2, \ldots.$$

This means that the probability of getting *x* flakes depends on the parameter λ (the mean, or expected number of flakes per 1 m^2 unit) and e (= 2.718...). x! ("x factorial") is the product of all the integers from 1 to *x*. If, for example, λ = 3, then the probability of finding two flakes in a unit is:

$$p(2) = \frac{e^{-3}3^2}{2!} = \frac{(2.718\ldots)^{-3} * 9}{2 * 1} = \frac{0.448}{2} = 0.224$$

Doing this for other values of x leads to the distribution in Fig. 8.5. Although this figure does not show bars for values greater than nine, that is only because their probabilities are so low that they are nearly invisible at the scale of the graph but, even for x = 100, they never quite reach zero.

If we were to ask how likely it was to find a single excavation unit with more than five flakes (H_A: p > 5), a glance at the graph suggests that this would be very unlikely *if flakes are distributed randomly on the site*. It is unlikely because the probability of this outcome is the difference between 1.0 and the sum of all the probabilities for outcomes of 0–5, (1–0.91608, or 0.08392, a fairly unlikely outcome). This brings out a useful aspect of the Poisson distribution; it is not that we think that the flakes actually will be randomly distributed; testing against a Poisson model gives us grounds for establishing that they are *not* randomly distributed. In fact, many cultural and natural processes result in the clustering of artifacts, and it stands to reason that we could measure the degree of clustering by how the actual distribution compares with the Poisson distribution. The simplest way to do this is with the variance-to-mean ratio; because, in Poisson distributions, λ is both the mean and the variance, this ratio should be very close to 1.0 for Poisson-distributed samples but will be noticeably greater than 1.0 for clustered samples.

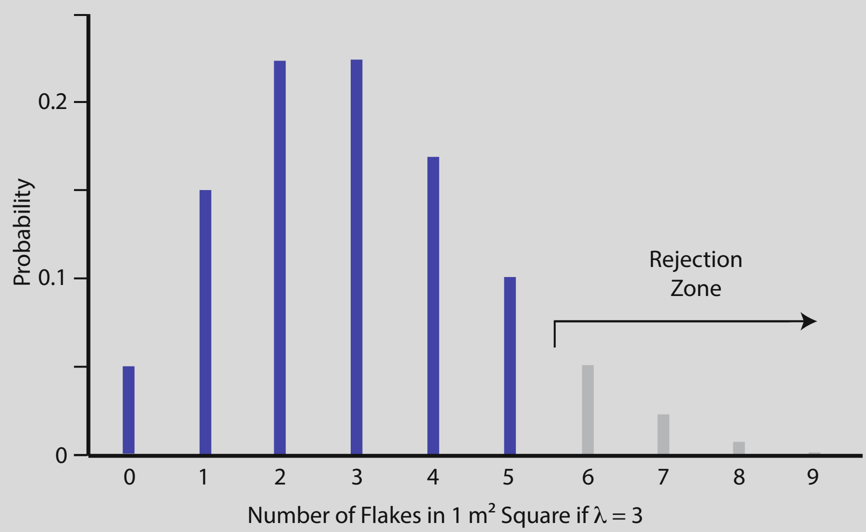


Fig. 8.5 Test of the hypothesis that the flake density could at a site could be λ = 3 if a sample drawn from the population has a mean > 5. This uses a one-tail test and the Poisson model and shows that this outcome has a probability of only 0.084 (grey bars), so that we would reject the hypothesis at 90% confidence but could not reject it at 95% confidence

Having said that, why should we care about confidence intervals? Even though they do not literally describe the range within which the true value lies, they do quantify the result of a measurement or statistical test in a way that allows us to assess the plausibility of an outcome. Even though the true value could lie outside our confidence interval for any individual sample, we can take some comfort in knowing that, 95% of the time, any random sample we take from the population will include the true value within its 95% confidence interval. The confidence interval also has a close relationship to tolerance (see Chap. 1), a key aspect of Quality Assurance and acceptability sampling.

A confidence region is the multivariate version of a confidence interval. For example, a bivariate normal distribution — perhaps for the probable location of an undiscovered site on a map — could have confidence intervals in both the North-South and East-West axes of the map, creating an elliptical confidence region.

Bayesian analysis (see below) has a similar concept called **credible interval**. However, it differs in some respects, including that the credible interval involves using a prior distribution to estimate prior probabilities, and that it treats the interval as fixed and the parameter as a random variable, rather than a "true" value (Jaynes 1976). However, the confidence interval and credible interval are sometimes equal.

8.6 Bayesian Inference

As already noted, not all archaeologists use classical null-hypothesis significance testing (NHST). An increasingly important alternative is Bayesian inference, which, rather than refer to the probability of the data given a null hypothesis, refers to the probability that a hypothesis is true given some data (Otárola-Castillo and Torquato 2018: 436). It is based on the multiplication rule of probability: p(A and B) = p(A) $*$ p(B|A). Rev. Thomas Bayes noted in a paper posthumously read at the Royal Society (Bayes 1763) that this relationship has very useful properties (Buck et al. 1996: 65; Salmon 1982: 51–55).

Bayes noticed that, because you can interchange A with B, the probability that both A and B will occur is.

$p(A\ and\ B) = p(A) * p(B|\ A) = p(B) * p(A|\ B)$, so, by rearranging terms,

$$p(A|B) = \frac{p(A) * p(B|A)}{p(B)}$$
$$= \frac{p(A) * p(B|A)}{p(A) * p(B|A) + (1 - p(A)) * (1 - p(B|A))}$$

This rearrangement of the multiplication rule is known as Bayes' theorem and provides the basis for a whole school of analysis. What it means is that we can evaluate the impact of a new piece of information, which in this case we call B, on the probability of some event A. p(A) is the prior probability of the event, as based on our original knowledge, and p(A|B) is the posterior probability in the light of evidence B. One of its advantages over NHST is that it does not call on us simply to reject or provisionally accept some hypothesis, it allows us to evaluate the relative merits of two or more competing hypotheses.

The thing that most worries archaeologists who are unfamiliar with Bayesian probability is the source of the prior, p(A). They may be tempted to use what are called "uninformative priors" (such as assigning all possible outcomes equal probability) but this does not make the test more objective, it is merely an admission of having no prior information. That is fine for cases where we really do not know anything, or for "nuisance priors" that are not very important to the main research questions, but one of the beauties of Bayesian probability is that it allows us to incorporate prior knowledge, whether that comes from quantitative analyses of pre-existing data or from more subjective expert opinion, in a formal way (Buck and Meson 2015: 571–573). That is a real advantage, as long as it is exploited carefully. One key feature is that, when using expert opinion to decide prior probabilities for a set of competing hypotheses, their values should reflect proportionality in the experts' prior beliefs; that is, if an expert assigns prior probabilities of 0.4 and 0.2 to hypotheses C and D, this should accurately reflect her belief that C is twice as likely as D (Frost 1999). "Knowledge elicitation" is the process of formalizing and calibrating these expert beliefs (O'Hagan et al. 2006).

In archaeology, Bayesian inference has had its greatest impact in chronology, and especially in the interpretation of stratigraphy and radiocarbon dates (see Chap. 21). However, it has considerable potential for broader application, including in spatial archaeology, in grouping artifact assemblages (Cullberg et al. 1975), and as an alternative to NHST inference.

8.7 Summary

- Most archaeological inferences are inductive because they are ampliative and often rely on probability
- "Verification" or testing of archaeological hypotheses is usually statistical, leading to a statement of the probability of obtaining our evidence if a hypothesis is true
- A common form of this evaluation is null-hypothesis significance testing (NHST), whereby we compare the statistics of a sample with a model for a null hypothesis, such as no difference between populations from which samples were drawn. This evaluates the probability of the evidence (sample characteristics) if the null hypothesis is true
- All such tests depend upon models, commonly including the binomial, Poisson, normal and t-distributions, although others are appropriate in specific situations
- An alternative to NHST is Bayesian evaluation of competing hypotheses. This evaluates the probability of a hypothesis, given some evidence, making it very useful when we want to compare competing hypotheses. Bayesian statistics are also the basis for some very important archaeological analyses, such as radiocarbon dating, but are useful in many other applications.

Case Study
Bayesian Inference

Otárola-Castillo and Torquato (2018) illustrate the use of Bayesian inference to choose among three hypotheses for two samples (dating early and late) of projectile points: H_1: that their average lengths fall within the range for a population of arrow tips, H_2: that their average lengths fall within the range for a population of *atlatl* dart points, and H_3: that their average lengths fall within the range for a population of spear points, with population means and standard errors for the three hypotheses of 6.9 ± 2 cm, 11 ± 2 cm and 14 ± 2 cm respectively. The samples had mean and standard deviation of 6.1 ± 2 (n = 10, early period) and 13 ± 3.2 cm (n = 9, late period). For demonstration purposes, they assume no prior information about the probabilities of each hypothesis (uninformative priors), thus making them all equal.

They computed the posterior probabilities by using Markov Chain Monte Carlo methods, which provide a way to estimate the shapes of posterior probability distributions by repeated sampling of probability density models. These kinds of models are the key to radiocarbon calibration and interpretation (Chap. 20).

The result, after very many iterations, showed, for the early period, a 0.97 probability that the projectile lengths were within the range for arrow tips, and only 0.004 and 0.00003 of the other two hypotheses. For the later period, the probability of H_1 was only 0.001, with 0.16 for H_2 and 0.89 for H_3.

The exact values of the posterior probabilities are not that important, as they depend on the priors selected, but some of the posterior probabilities are so low that we can reject them with considerable confidence. Meanwhile, the posterior probability for H_1 is very high in the early sample, providing grounds for confidence that the early projectile points were for arrows. For the later period, the probability of almost 0.9 that the projectiles are in the length range for spear points is also fairly compelling and indicates that H_3 is much more probable than H_2.

Notably, the Bayesian analysis allows transparent evaluation of the probabilities of the three hypotheses, something that NHST does not do.

References Cited

Baxter, M. J. (1994). *Exploratory multivariate analysis in archaeology*. Edinburgh: Edinburgh University Press.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society, 53*, 370–418.

Buck, C. E., & Meson, B. (2015). On being a good Bayesian. *World Archaeology, 47*(4), 567–584. https://doi.org/10.1080/00438243.2015.053977.

Buck, C. E., Cavanagh, W. G., & Litton, C. D. (1996). *Bayesian approach to interpreting archaeological data*. New York: John Wiley & Sons.

Cullberg, C., Odén, A., Pehrsson, N. G., & Sandberg, B. (1975). Some quantitative methods applied to flake-measuring problems. *Norwegian Archaeological Review, 8*, 81–97.

Doran, J. E., & Hodson, F. R. (1975). *Mathematics and computers in archaeology*. Edinburgh: Edinburgh University Press.

Drennan, R. D. (2010). *Statistics for archaeologists. A commonsense approach*. New York: Springer.

Fletcher, M., & Lock, G. R. (1991). *Digging numbers: Elementary statistics for archaeologists*. Oxford: Oxford University Committee for Archaeology.

Frost, J. R. (1999). *Principles of search theory part III: Probability density distributions*. Fairfax: Soza & Company, Ltd.

Jaynes, E. T. (1976). Confidence intervals vs. Bayesian intervals. In W. L. Harper & C. A. Hooker (Eds.), *Foundations of probability theory, statistical inference, and statistical theories of science* (pp. 175–257). New York: Springer.

Kadane, J. B. (1988). Possible statistical contributions to paleoethnobotany. In C. Hastorf & V. Popper (Eds.), *Current paleoethnobotany. Analytical methods and cultural interpretations of archaeological plant remains* (pp. 206–214). Chicago: University of Chicago Press.

Krakker, J. J., Shott, M. J., & Welch, P. D. (1983). Design and evaluation of shovel-test sampling in regional archaeological survey. *Journal of Field Archaeology, 10*(4), 469–480.

Lee, G. A. (2012). Taphonomy and sample size estimation in paleoethnobotany. *Journal of Archaeological Science, 39*(3), 648–655.

Morey, R. D., Hoekstra, R., Rouder, J. N., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin and Review, 23*(1), 103–123.

Neyman, J. (1937). Outline of theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A, 236*(767), 333–380.

O'Hagen, A., Buck, C. E., Daneskhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., et al. (2006). *Uncertain judgements: Eliciting experts' probabilities*. London: Wiley.

Otárola-Castillo, E., & Torquato, M. G. (2018). Bayesian statistics in archaeology. *Annual Review of Anthropology, 47*, 435–453.

Rowntree, D. (2018). *Statistics without tears: A primer for non-mathematicians*. London: Penguin Books.

Salmon, M. (1982). *Philosophy and archaeology*. New York: Academic.

Shennan, S. (1997). *Quantifying archaeology*. Edinburgh: Edinburgh University Press.

9 Basic Artifact Conservation and Lab Management

... archaeological artifact conservation starts well before any artifact is excavated and continues well after an artifact is curated

(Rodgers 2004: 8).

Most kinds of artifact conservation require a strong background in chemistry and materials science, and there are no simple "recipes" for the conservation of objects (Cronyn 1990: xv). This chapter focuses on basic care and handling of artifacts, storage and monitoring of collections, and lab management. The only remedial action discussed here is the most basic cleaning, and refitting of fragmented pottery. For further reading on conservation, including broader issues in conservation, readers should consult the specialized literature (e.g., Applebaum 2007, 2018; Bradley 1990; Fink 2017; Rodgers 2004; Sullivan and Childs 2003; UKIC 1983).

The purpose of artifact conservation is to preserve and maintain the stability of artifacts and collections while intervening as little as possible with the archaeological evidence (Cronyn 1990: 9; Richmond and Bracker 2009). Ensuring an environment that does not accelerate their deterioration is one of its most basic tasks. More aggressive treatments, such as removal of corrosion on metal or applying consolidants to prevent disintegration, may make the artifact useless for some kinds of archaeological and archaeometric analyses, and more passive methods are preferable where possible, but the ethics of non-intervention can be complicated (e.g., Kemp 2009; Viñas 2009).

Another important conservation ethic is reversibility. Although no treatment is completely reversible, conservators aim to select treatments that cause as little permanent alteration to the nature of the piece as possible.

A third important ethic is to maintain accurate records of any actions taken to clean or preserve the artifacts, as well as of any observations about the artifact that might be important, and especially observations of things that might not survive or need to be removed. Documentation ensures that a future conservator knows what procedure to use should it be necessary to remove adhesives or consolidants applied previously.

Fourth, artifact conservation should be collaborative (Cronyn 1990: 10). Conservators work with archaeologists, museum curators and others to determine the best way to protect the long-term integrity of objects as part of a collection management strategy (Childs and Benden 2017).

9.1 Artifacts in the Burial Environment

While artifacts are surrounded by sediment or water, or even while they are still in use, various physical, biological, and chemical processes begin to affect their condition (Cronyn 1990: 14–29; Dowman 1970: 4–47). Handling during use (Skibo 1992) and movement of particles in sediments alters the artifacts' surfaces. Percolation of water through sediment and alternation of wet and dry conditions, for example by groundwater fluctuations, can dissolve some chemicals and remove them or deposit them by precipitation, leading in some cases to the mineralization, or fossilization, of wood and bone, corrosion of metals, accumulation of salts in or growth of thick carbonate coatings on pottery and bone. Consistently wet or very dry conditions, or presence of copper, may allow organic materials to survive for extremely long periods, albeit in altered form, by precluding bacterial action. The wooden members of sunken ships, for example, may perfectly preserve their form even though the contents of the cells in the wood have been replaced by sea water (Cronyn 1990: 26; Jespersen 1985; McCawley 1977; Pearson 1981; Unger et al. 2001). The environment's pH is very important (Dowman 1970: 18–28) because inorganic

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_9

materials, especially metal, often deteriorate in acid environments, while organic materials are often degraded in basic environments. Bone, antler, and ivory have both organic and inorganic components; thus, in basic environments, bone may have structural integrity but contain no organic collagen useful for things like radiocarbon dating (Bradley 1992). Salinity, temperature and organisms also affect materials (Kibblewhite et al. 2015).

Although many destructive processes are slow and continue indefinitely, quite often artifacts, after a brief, initial period of rapid deterioration, reach equilibrium with their environments and stabilize in a state that changes very little (Cronyn 1990: 29; Dowman 1970: 4). However, removal of an artifact from its environment can restart the process of rapid destruction. Consequently, it is often better to leave it where it is or keep it in conditions that mimic the burial environment until it is possible to stabilize it at a new equilibrium (UKIC Archaeology Section 1983).

We improve artifacts' chances of survival by making their transition from the burial environment to a lab or museum environment as gradual as possible. This can include preventing them from drying out too quickly, using containers that "breathe" to prevent condensation, and avoiding sudden exposure to sunlight.

Some kinds of artifacts and samples need particularly tender treatment. Very fragile ones might need to be removed to a conservation lab while still enveloped in sediment, as was the case with the Neolithic plaster statues of 'Ain Ghazal (Tubb 2001). More typically, this involves physically supporting the sample with gauze strips and a stiff under-support. Samples that will be subject to chemical analysis, radiometric dating, or DNA analysis also require special care to avoid contamination. This can include careful removal of sample material from surrounding sediment with clean tweezers or collection of sediment from nearby locations as a check on contamination or evidence of radiation background.

9.2 Handling Artifacts in the Laboratory

Artifacts are often fragile, and sometimes need to be protected from contamination, so the way we handle them is very important (Bradley et al. 1990; Meister 2019: 268–269; Rodgers 2004: 41).

One of the greatest risks from handling is breakage. Before you pick up an artifact, you should assess its probable strength and vulnerabilities. Do not grasp it by projecting parts, such as handles, as these might break off. Instead, try to support it under its center of gravity, preferably with something that cushions it (Miles 1992). The most fragile ones should not be held in your hands at all, but on some kind of support, such as a tray with gauze or soft, stable foam padding to support weak points or irregularities of shape, but not stick to or catch on edges of the object.

Even relatively stable and sturdy artifacts are susceptible to breakage if dropped on a hard surface. Despite your best efforts to handle artifacts carefully, there is always a chance that you will drop one eventually, so it is essential to hold artifacts over a cushioned surface that will absorb much of the shock of impact. Cover laboratory surfaces with sheets of acid-free padding, such as Nalgene, polyethylene, Styrofoam, bubble pack, or carpeting. Ideally, use a material that lies flat, with minimal curling of edges, and that will not slide on the lab counter. Not all such padding would be appropriate for use in long-term storage (see below).

Some kinds of artifacts should never be handled with bare hands. The skin in your hands leaves oils and organic acids on surfaces that you touch. These can actually etch into metal surfaces slightly, provide sites for future corrosion, and contaminate artifacts with substances that react chemically with the artifacts' material, especially metals (Shearman 1990). Sometimes these contaminants could also confuse the results of archaeometric analysis or radiometric dating. In any of these cases, you should handle the artifact or sample with clean tools, such as tweezers made of an inert material, or wear nitrile gloves. Nitrile gloves are available in boxes of 100 or more pairs and should be present in all archaeological laboratories. Cotton gloves are no longer recommended because sweat and salt from your skin can wick through the fabric onto the artifact. Even while using gloves, you should minimize wear on the surfaces of metal artifacts, especially coins, by holding them only by their edges, and change gloves periodically to avoid transmitting dirt from one artifact to another. You should also avoid wearing jewelry if you are handling artifacts.

If you have to transport artifacts more than a very short distance, or the artifacts are heavy, carry them in a tray or box or on a wheeled cart. Boxed artifacts should have a packing list, and artifacts in the box should not be wrapped in paper that might be thrown away, as this poses a risk that small objects could accidentally be overlooked and discarded. Checking the contents against the packing list helps to avoid this.

9.3 Simple Cleaning of Artifacts

Some kinds of artifacts, typically lithics and well-fired pottery fragments from moderate burial environments, are sufficiently stable that archaeologists can clean them without the help of a professional conservator (Rodgers 2004: 43–44). However, you should always examine artifacts carefully before cleaning in case cleaning might remove important evidence, such as pigments, slips, use residues, fibers or other organic materials. It is also a good idea to test a few artifacts to make sure they are as stable as you think. If there is any risk that cleaning might damage or break the artifact, you should draw or photograph it and make some basic measurements before proceeding.

As long as the artifacts will not be used for a kind of analysis that cleaning or chemical contamination would compromise, and they have hard surfaces, you can remove most loose dirt by brushing with a soft brush and remove tougher dirt mechanically with a fine tool, such as a scalpel, bamboo stick, or dental pick. You should use brushes that suit the situation — soft ones for loose dirt or vulnerable surfaces — and never use metal brushes, vigorous brushing, or scraping, as these will damage the surface of pottery and the edges of lithics. If you still find it difficult to avoid damage to delicate slips on pottery, you should record any damage you have done and make any observations on the slip that you still can (see pp. 194–195).

When the adhering dirt is very stubborn and the artifacts are able to withstand immersion in water and have not absorbed a lot of salt (test one or two first), you may use water to soften and loosen the dirt before and during brushing and mechanical removal such as lifting crusts with a small tool. This may still not be enough to remove very hard carbonate deposits, making it necessary to use a relatively gentle acid, such as acetic acid (vinegar) or dilute hydrocholoric acid. Soak the artifacts first, to prevent the acid from penetrating into the artifact's pores, and it is generally better to drip the dilute acid onto stubborn precipitates with an eye-dropper than to immerse the artifact in acid, as this provides better control.

When using any acids, you should consider the safety protocols (see Chap. 10) and seek the advice of a conservator to ensure that the acids you use will not damage the artifacts, which often have components that are chemically similar to the material you are trying to remove. In addition, you should not use organic acids such as vinegar on any artifact that might contain lead, including lead pigments on pottery, as that would cause rapid corrosion. Finally, it is important to neutralize and remove all traces of the acid when you are finished, generally by rinsing in distilled water, to prevent the acid from continuing to act on the artifact.

9.4 Storage of Archaeological Collections

Archaeologists and curators must take reasonable steps to preserve both the physical and archaeological integrity of the artifacts under their care. This means protecting them from environmental conditions that could accelerate their physical deterioration or result in loss of archaeological, including contextual, information (Bradley and Daniels 1990; Leigh 1982; Meister 2019; Partington-Omar and White 1981; Pye 1992; Tate and Skinner 1992; UKIC Archaeology Section 1983, 1984).

Labelling The integrity of the contextual information for each artifact is at least as important as the physical integrity of the artifact itself. A basic principle is that no artifact should be separated from its labelled bag unless the artifact itself is labelled with a catalogue number that allows us to connect it with its archaeological context. Typically, redundancy in labelling (e.g., labelling the artifact, the bag it is in, and the drawer that contains the bag) helps to protect against failure in the recording system (Bleed and Nickel 1989).

Somewhat ironically, it is often the most remarkable artifact whose exact context is lost, simply because someone extracted it from other artifacts for drawing, photography, conservation, or to show colleagues, before it was properly labelled.

Keeping unlabelled artifacts in their own labelled bags is not the best solution, since artifacts can always be removed from their bag, or be replaced in the wrong one. More usual is a number inked right onto the artifact's surface and it should be legible, reliable and stable (Pye 1992: 398). Obviously, we would not want labels to fall off accidentally while in storage, or to be too difficult to read. This need to protect contextual information permanently creates tension with the conservation principle of reversibility.

Consequently, we ensure that every artifact receives a label that is reasonably permanent without intentional intervention but can be removed by a conservator (reversible). For lithics and pottery, this typically involves painting a small strip on the surface of the artifact, in a spot that is unobtrusive but visible, using an acrylic, such as Paraloid or Acryloid B72, dissolved in acetone (Koob 1981, 1986). Once this hardens, it provides a clean, smooth surface on which to write in India ink with a fine technical pen or affix a small printed label, or even a bar code. We then seal the label with another layer of the acrylic. The ink we use must be one that does not dissolve in acetone but, if we need to remove the label, we can use acetone to dissolve the layer of acetate, so that the whole label comes off cleanly. This kind of label is fairly stable but can still detach under poor environmental conditions (see Sect. 9.5) or if the artifact was not sufficiently dry when labelled. Water in the pores of a sherd, for example, can prevent the acrylic from adhering properly.

Storage Artifacts' storage conditions should minimize the rate of processes that deteriorate them. This involves maintaining a moderate environment that is stable in temperature and humidity and that protects artifacts from sudden movement, friction and vibration. The storage area should be dry, clean, secure, away from traffic and heating systems, but accessible with wheeled carts for moving trays and boxes (Pye 1992: 400–401; Rose and Hawks 1995; Tate and Skinner 1992). It should be away from direct sunlight, and fluorescent lights should have UV filters or be left off in "dead storage" unless the storage cabinets prevent light penetration.

High temperatures accelerate chemical reactions, but extremes of heat and cold are avoidable by curating artifacts in a climate-controlled building with temperatures close to 20 °C.

Humidity is particularly damaging to metal artifacts (Clarke and Bradshaw 1982; Knight 1982; Shearman 1990) and ceramics infused with soluble salts from groundwater. A simple way to prevent humidity from climbing too high is to put artifacts in a sealed polyethylene box, creating a "micro-climate," along with silica-gel, which absorbs moisture from the air in the container into its micropores until it comes into equilibrium with its surrounding air (Fig. 9.1). The silica gel thus becomes saturated after some months, and no longer works, so you must revitalize it by heating in a lab oven at 105 °C for 10 h. This drives off the accumulated water molecules. Let it cool for another 10 h and then replace it in the box. To signal when the silica gel needs renewal, you can either put a humidity strip — a small paper strip that indicates by color changes the relative humidity — in the box or use an "indicator" silica-gel that itself changes color when it needs changing. A hygrometer in the container allows you to monitor humidity even more precisely. For very vulnerable artifacts, such as iron ones, you may need enough silica-gel in the container, and change it often enough, to keep relative humidity between 10 and 20% (Cronyn 1990: 75). A recommended volume is 1 kg of silica-gel for each 0.012 m^3 of air to be desiccated (Watkinson and Neal 1998). Because the effectiveness of silica-gel is related to surface area, it is often better to spread a layer of silica-gel across the bottom of the box than to leave it in a small pouch, although this has safety implications and you should consult the hazard description sheet for silica gel (p. 153).

Biological hazards to artifacts, such as insect or rodent infestation or mold and mildew (Stansfield 1985) are most critical for organic materials, such as wood, bone and antler. Use of storage systems that exclude vermin, keeping the lab clean and preventing anyone from eating or drinking in the lab are preventative measures.

Apart from environmental conditions, the physical means for containing and organizing artifacts is the most important storage factor. The storage system should organize the artifacts in a way that facilitates their location and retrieval, allows easy and safe removal and replacement of artifacts, minimizes their exposure to movement, friction and abrasion, facilitates monitoring, and maintains the collection's environment within the safe zone. We need to consider the materials with which the storage system is constructed, the shape and size of units within it, how robust and chemically active the artifacts are, and how and how often we need to examine or remove artifacts. We may consider various types of shelving and cabinetry, drawers, trays, boxes and bags. The preferred storage solutions are constructed from inert materials that will not release harmful acids or other chemically active agents into the artifacts' environment.

The storage system should make sense archaeologically, facilitating the collection's use. Separating artifacts by categories of varying fragility or vulnerability and with different environmental needs is also important (Cronyn 1990: 79–80; UKIC Archaeology Section 1982). Category A is for the most unstable materials, for which control and monitoring of the environment is most important. These artifacts, often

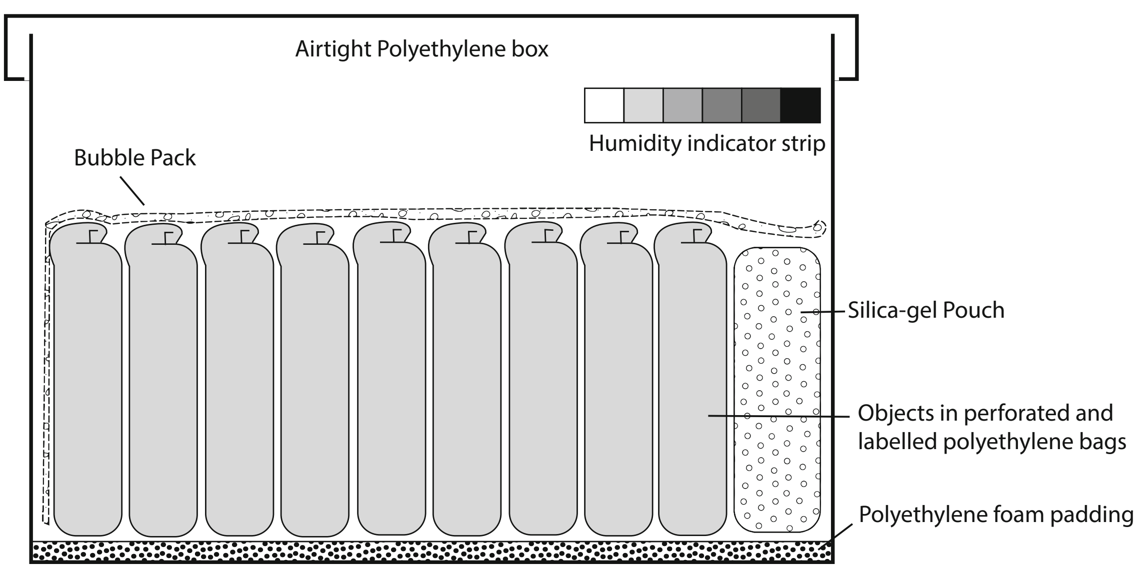


Fig. 9.1 Schematic of a storage box containing silica gel to control humidity for sensitive artifacts, such as metal ones (W. Wadsworth, after Cronyn 1990: 75). Note that it would be necessary to change the silica gel periodically, on the basis of the humidity indicator strip

of metal or organic materials, need enclosed containers with sealed and carefully controlled environments. Category B refers to less sensitive objects, such as bone, that still need to be protected from damp, fungus, pests, and abrasion. Category C is for relatively robust objects that have greater tolerance of environmental variation, such as lithics and most kinds of well-fired pottery. Even for Category C, relative humidity should not exceed 65% to prevent mildew on artifacts, labels, or packing materials.

A computer database that records the storage location of each object by its catalog number makes it easy to find things from any archaeological context and to track artifacts moved or removed for various purposes (see Sect. 9.9).

A common storage system for Category C consists of sturdy wooden shelving to support wooden or acid-free cardboard boxes and trays in which artifacts are spread out, piled, or arranged in rows or bags. The shelving is in rows in a dry room with a stable temperature around 20 °C. The boxes or trays are arranged in a sensible order, each with a label on the end facing out to summarize its contents. This system is useful for "dead storage" of artifacts not currently under analysis but has major disadvantages that make it inappropriate for more vulnerable artifacts or active analysis. The wooden structure, typically of softwood, contains resins and exudes acidic vapors. Plywood and particle-board, often used in the shelving, contain glues that may also give off harmful organic acids. Cardboard boxes are susceptible to water damage, are usually made from acidic paper and glue, and do not protect their contents very well. Friction and jostling of artifacts as boxes are moved causes abrasion and encourages breakage. If the boxes or trays are heavy, they are difficult to move, pose potential injury to lab workers, and moving them increases opportunities for breakage. We can compensate for some of the problems by covering surfaces with aluminum foil barriers or substituting metal shelving and acid-free cardboard or plastic boxes for wooden ones, but these are still inadvisable for anything but dead storage of relatively robust objects.

A better storage solution where it is necessary to retrieve and replace artifacts frequently is one with drawers that minimize friction when pulled out, and that are lighter and less cumbersome than heavy wooden trays. An excellent but expensive storage system for this situation consists of metal cabinets that seal when closed, and that have many shallow, metal trays as drawers (Fig. 10.1). This isolates sets of artifacts in a sealed environment enclosed by inert materials, which may help maintain constant and low humidity. The shallow trays on adjustable tracks can be lined with chemically inert padding, such as polyethylene (Pye 1992: 396), and arranged to accommodate either large or small artifacts, while subdividing the trays with strips of inert material organizes the artifacts into rows. The tray fronts should have slots for labels to identify drawer contents.

A similar but less expensive alternative employs many tray-like plastic boxes that either slide along closely-spaced metal shelves or have flanged edges that slide along tracks. Even better than the metal cabinets, the plastic boxes provide sealed environments that, with silica-gel placed in each, are particularly suitable for metal and some other vulnerable artifacts of Categories A and B. Plastic boxes should best have transparent lids that allow monitoring of humidity indicators without having to open them.

Even in well-designed systems, every time someone pulls out or replaces a drawer, its contents may shift slightly, causing wear through abrasion. Lining drawers or artifact containers with a material that prevents them from moving or cushions them from hard surfaces can reduce this effect. Even the simple expediency of putting artifacts in individual polyethylene bags protects them somewhat. For artifacts that are particularly vulnerable to abrasion, cutting form-fitting holes in a polyethylene pad that fits into the tray can be used to hold each artifact in place and prevent shifting when the drawer is moved. The key is to have a system that prevents artifacts from jostling around yet allows easy retrieval so that there is less risk of damaging pieces while they are being removed or replaced.

9.5 Collection Monitoring and Collection Census

Even nearly ideal conditions do not guarantee that artifacts will not deteriorate. Thus, it is important to make regular, preferably scheduled, checks on collections' condition, especially for Categories A and B, replace silica-gel as needed, record any changes that have occurred and check on environmental conditions. This is called **collection monitoring** (Bradley and Daniels 1990; Keene and Orton 1991) and the records of the artifacts' conditions constitute a condition census.

9.6 Conservation Documentation

It is important to record information about an artifact both before and during any conservation treatment it receives, as well as to record the treatments themselves and the artifact's storage history (Bradley 1983; Corfield 1983; Rodgers 2004: 15).

In addition to the usual information that an archaeologist would want to record about the artifact, these records include a brief description of the artifact and its condition before treatment, photographs, details of how the conservator examined the artifact and what that revealed, and a step-by-step description of any treatment it received. The treatment record should include details of investigative cleaning, methods of stabilization and materials used, and the resulting condition. It may include recommendations for further treatment, if needed, and requirements for its storage environment (Cronyn 1990: 94–95).

9.7 Refitting Pottery

One of the few remedial measures that archaeologists routinely carry out without necessarily involving a professional conservator is refitting lithic flakes to reconstruct cores or reassembling fragmented pottery to restore vessels. The former tends to be temporary, often to determine the order of flake removal, while the latter may be intended to be more permanent. Refitting can also be an important aspect of studying site-formation processes (see pp. 169–170).

For pottery, there is more to refitting than just trial and error, and you should plan the job after considering several questions (Cronyn 1990: 157). First, is it really necessary to join the pieces physically, with adhesives? Or do you only need to discover which pieces fit together and reconstruct the vessel virtually on a computer? Is the condition of the pottery good enough to withstand reconstruction? Flaking or friable sherds as well as porcelain and stoneware are not good candidates for physical reconstruction. Are the breaks free of dirt and encrustations of salt or carbonates? If not, they will require additional cleaning before proceeding with reconstruction. Are the sherds sufficiently dry? Moisture in sherds prevents consolidants from penetrating properly. What previous conservation treatments have the sherds undergone, if any?

Once you have answered these questions and are prepared to go ahead, you start by laying out the sherds with either the outer or inner surface up. Close attention to the shapes, colors, fabric, coil or wheel-throwing traces, and any decoration that might be present on sherds will help to determine the sherds' orientations, whether or not they are likely to come from the same vessel, and sometimes their approximate position on the vessel. For example, you will want to remove sherds whose fabric or thickness indicates that they could not be from the same vessel as most of the sherds. Then group rim sherds in one place, base sherds in another, and, if possible make other groups for neck, shoulder and body (see pp. 246–247 for parts of vessels). Try to orient the sherds in each group the same way (generally with their likely top away from you). Rills from wheel throwing, coils, and some kinds of decoration that you would expect to be horizontal help you orient sherds correctly. Sometimes you can expect thicker body sherds to be closer to the base. Then you use the sherds' shapes along with other clues, such as rilling, curvature or decoration, to help you find sherds that may fit together. Test potential refits by gently fitting them together, without adhesive, to see if they fit snugly. Sherd color can be deceptive, as sherds that fit together sometimes differ considerably in color because they experienced different burial environments, or because one was burned after breakage.

Once you have identified sherds that fit together, record which these are. This not only makes it unnecessary to go back over old ground if your refitting is interrupted, but is important for understanding site-formation processes. If sherds from the same vessel were only found in one context, this would suggest, among other things, that the context may not have been disturbed substantially. Finding joins between sherds from many contexts, by contrast, tells you that the sherds were scattered after breakage, and patterns in their distribution may indicate which contexts were likely contemporary or which activities moved the sherds around.

If you decide to rebuild the vessel physically, you should plan your reconstruction carefully, having already discovered all the joins and ensured that sherd edges are clean and dry (Cronyn 1990: 158–159). Plan the order in which you will physically join the sherds and try to rebuild the vessel in one session, if possible. This helps to avoid "locking out" sherds that cannot be fit between previously fitted sherds, or failure to match up different parts of the vessel as you work your way around because of small errors in curvature while you work. For whole vessels, it is advisable to start with the base and build upwards, or at the rim and build the vessel upside-down to ensure that the curvature for the whole circumference is correct.

The three ethics outlined at the beginning of this chapter dictate that the adhesive we use should be one that causes minimal alteration to the chemical and physical nature of the sherds, that we can easily remove it, and that we record what kind of adhesive it was (Down 2015). Usually, Paraloid or Acryloid B72 dissolved in acetone is a good choice. It holds sherds together well, once the acetone evaporates, but the evaporation of acetone calls for use of a fume hood to protect lab workers (Chap. 10). Joins made with B72 are reversible by re-introducing acetone, which allows you to correct errors. Non-reversible adhesives make it difficult to correct such errors without risk of breaking the sherds. One adhesive that has been very popular among archaeologists, "white glue" (a collagen-based adhesive), can be removed but does not hold the pot's shape under hot or humid conditions. It also contaminates sherds with collagen, thus making them unsuitable for residue analysis or radiocarbon dating.

Priming the edges of the sherds with a 5% solution of B72 can be helpful. Once this has dried, apply 25% B72 in a thin trail along the edge of one sherd, and then press the sherds snugly together (Koob 1986). You may have to rock the sherds back and forth slightly to get a tight join. After holding the sherds firmly together for a few minutes, stand them in a "sandbox" (filled with dried beans or plastic beads, rather than sand that might get stuck in crevices). Some people also reinforce the joins temporarily with tape, but you should ensure that the tape will not damage slips or leave unwanted glue. After a short period, begin to add further sherds in the same way. Resist the temptation to join disconnected parts of the vessel at the same time. While that might seem more efficient, it poses the risk that small differences in curvature would later make it difficult or impossible to fit the large portions together, so you need to remove or loosen the joins with acetone and start over. Instead, build upwards from the base or rim, preferably in a spiraling sequence.

Note that if you use dried peas or beans for your "sandbox," you should ensure it is sealed with a tight lid while not in use. Otherwise, it will attract pests, such as mice.

9.8 Removing Samples for Analysis

Although our usual concern is the stabilization and protection of the artifacts, some kinds of analysis are invasive or destructive. For example, petrographic analysis of pottery or lithics requires thin-sections made by polishing slices sawn from artifacts. Many kinds of chemical and isotopic analysis and dating methods also require us to remove at least small amounts of material from artifacts, bones or teeth. Clearly, it is important to minimize the extent of destruction while also maximizing the effectiveness of the analysis, which calls for balance and attention to both the artifact's characteristics and the questions archaeologists are asking. For pottery, for example, we might select cutting or drilling locations that minimize damage and avoid decorative features (unless it is a decorative paint we wish to analyze), or we might select a slice for thin-sectioning that gives us the maximum information for the amount of damage we inflict.

For thin-sections of pottery, we usually select a place that will provide a good radial section (see Figs. 16.10, 21.12) with the longest "height" or intercept important structural features (e.g., midway through a handle), which has the secondary effect of making the sherd easier to draw (see p. 358) or features like slips and sherd "cores" easier to measure (see pp. 200–202). For thin-sections of obsidian flakes, we would want to make a cut perpendicular to the surface so that it yields accurate measurement of hydration rinds for dating purposes (Shackley 1998).

Many kinds of archaeometric analysis, such as Instrumental Neutron Activation (INAA) and X-ray Diffraction, require powdered samples removed from artifacts by drilling. It is important to avoid contamination by using a clean drill bit made of a material that does not interfere with measurement of the constituents of interest, to discard the first material drilled out, in case it is contaminated by surface contact, and to place the drilled material directly into a clean, inert vial. A single drilling site may not characterize the whole artifact's composition well, so it may be advisable to take multiple sample elements from the same artifact. If possible, we choose unobtrusive drilling sites that will not obscure important features of the artifact, and always document the locations of removals.

Removal of material from bones or teeth for isotopic or ancient DNA research sometimes has ethical considerations, especially when from human remains (e.g., Prendergast and Sawchuk 2018).

9.9 Laboratory Protocols and Quality Assurance

Lab managers are not only responsible for the protection and curation of artifacts, they also need to ensure the quality of the research in the lab and that the lab meets safety and other guidelines and legal requirements. Among the ways they do these things are training of lab personnel and clear policies and protocols for lab activities, often either posted on a wall or bulletin board or kept in one or more binders in the workspace. Some examples of these follow.

Safety Protocols Health and Safety is extremely important and information should be clearly posted in the lab to indicate the location of a First Aid Kit and, in labs that use chemicals or hazardous equipment, the spill kits, data sheets for each type of chemical, shutoff switch, lab shower, eyewash station, etc. (see Chap. 10). A typical very basic protocol is prohibition of consuming food or drink in the lab.

Specimen Removed Log or Artifact Location Catalog Usually artifacts and other archaeological collections are stored in an ordered manner in cabinets, but it is common for them to be moved to some other location for photography, exhibit, special off-site analysis, or teaching. This poses a risk that items will be lost, so it is essential to record details of the loan or removal in a log-book or database, by tags placed directly in the storage shelves or drawers, or both (examples in Fig. 9.2).

Foreign Sediments Log Projects that import soils and sediments from outside the country normally need to have an import permit, and the terms of this permit include protocols for ensuring that these soils will not pose any danger to people or especially agriculture in the host country. Specifically, their provisions do not allow disposal in normal trash unless the samples have first been sterilized by heating in a lab oven for a period of time above a minimum temperature. Such protocols usually require a log that documents which sediments entered the lab when, where they are, and whether they have been destroyed or sterilized. They also usually require each container of sediment to carry a warning

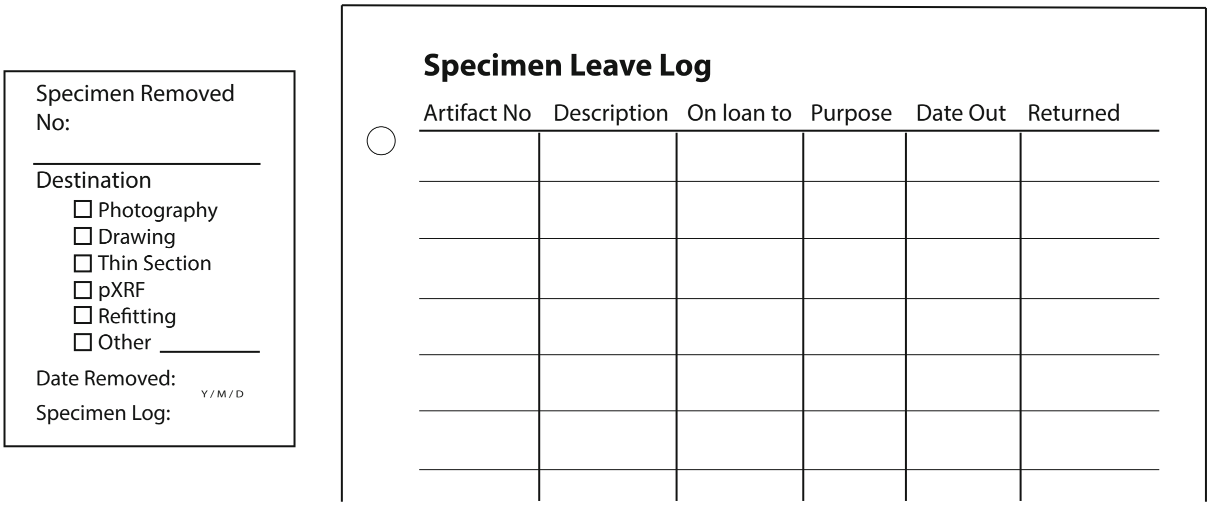


Fig. 9.2 Example of a tag to indicate the removal of an artifact from its place in a drawer (left) and part of a page from a "Specimens Removed Log" (right) that documents the date and purpose of removal and the person responsible

label, and often require a dedicated cleanup kit for spills. In some jurisdictions, there may be similar restrictions for imported faunal remains.

Registered Specimens Log This is just a record, either in a log book or in a database, that keeps track of the numbers used to catalog artifacts, essentially to ensure that each one has a unique number. In some instances, it may be combined with an Artifact Location Catalog.

Conservation Log This documents the treatments that individual artifacts received from a conservator, including stabilization treatments, chemicals used, repairs, and joins, and the dates of those treatments. It can also include drawings, photographs, or measurements that the conservator made prior to carrying out the treatment.

Measurement Protocols For each type of artifact or other item under study in a lab, there should be a protocol that specifies how artifacts should be classified or measured, including clear definitions or illustrations of concepts like "axial length" or "maximum dimension" for flakes, proportion of preservation for bones or vessel rims, roundness or angularity in ceramic inclusions, and so on. Classification protocols would normally have a taxonomic decision-tree to help people assign artifacts to a category unambiguously and consistently (see Chap. 3).

Illustration Protocols As discussed in greater detail in Chap. 21, archaeological illustrations are coded representations that require a very consistent set of conventions within a given project. Protocols for this would include line thicknesses, orientation of artifacts, and conventions (hatching, etc.) for representing colors, textures, polish, or surface treatments for pottery, lithics, metal and bone artifacts, and so on.

Data Protection Protocols Since modern archaeology uses computers to archive most of data, including fieldnotes, photographs, artifact descriptions, and stratigraphic records, it is necessary to ensure data security and preservation (Eiteljorg 2004; Kintigh and Altschul 2010). A minimal requirement is regular backup, but the fact that computer hardware and software keep evolving also means that file formats and computer hardware change, so it may be necessary to "migrate" files to newer file formats or media every few years. Another problem is that some storage media, especially older magnetic media, degrade rapidly over time. This corrupts the files so that they become unreadable. Use of cloud storage can mitigate some of these concerns, but does not eliminate them.

A particularly vulnerable resource today is any archive of digital photographs and even prints made from them. Although digital photography allows us to take hundreds of potentially useful images cheaply, and keep them in cloud storage, we need to be mindful that these can easily be lost, either for one of the reasons just mentioned, or simply because they have arbitrary file names that do not identify the image in a useful way. It is best practice to make hard copies of the most important images on archival-quality photographic paper or even glass negatives that will survive for many decades or even centuries if they are stored safely. Notably, prints made from digital images vary substantially in their image quality and stability (Knoll and Carver-Kubik 2019). A key ingredient of any photographic archive is a log or database that properly captions each photo so that we know what it is showing and can easily search for images we need.

But all digital data, not just photographs, are vulnerable. A potentially useful yet simple lab tool is a log, perhaps posted on the wall near a workstation, that records when each computer was last backed up. For the longer term, we need something more, that will also "migrate" the files as technology changes. Important institutions in this area are organizations like Digital Antiquity. It hosts the digital archive, **tDAR** with the objective to ensure long-term preservation of digital archaeological archives (https://core.tdar.org/; Sheehan 2015).

9.10 Ethical Issues

Curation of specimens not only involves their preservation; there are also broader ethical issues. Today, the most noteworthy issue involves the curation, use, and potential repatriation of artifacts or human remains claimed by indigenous peoples. It is also important to ensure that collections meet the conditions of the authorities that issued excavation and survey permits, import or export permits, or of funding agencies, and the laws of the country or state of origin and such international conventions as the UNESCO convention.

Any artifacts or specimens that may be culturally sensitive should be treated with respect and there should be protocols for both the handling and storage, and the legitimate repatriation, of items that are associated with or culturally important to extant indigenous groups (Clavir and Moses 2018; Sullivan and Edwards 2004). These typically include restrictions on the display of human remains or sacred artifacts, consultation with indigenous representatives, and sometimes provisions for the use of cultural objects by descendent groups in their ceremonies. In addition, the recent explosion of archaeogenetic (aDNA) research raises ethical issues about how or whether genetic material should be used or retained from repatriated remains (e.g., Prendergast and Sawchuk 2018).

More generally, archaeological permits usually have provisions for the ultimate disposition of artifacts from excavation and survey that can specify storage locations, place limits on the export or the duration of loans of artifacts, and, in some cases, even specify ways to dispose of de-accessioned artifacts that are not considered to have further research value. A key feature of most such policies is that artifacts should not be commercialized. Import permits for foreign soils, sediments and plant remains will also specify the exact conditions under which they should be stored and how to dispose of them when research is complete, typically by incineration or sterilization in lab ovens.

9.11 Summary

- Artifact conservation involves stabilizing them with as little interference as possible and maintaining them in a stable environment that discourages deterioration
- Preservation of artifact documentation is just as important as preserving the artifacts themselves
- Removal from the burial environment often accelerates deterioration
- It is important to handle artifacts carefully, in ways that prevent damage or cross-contamination, and only over padded surfaces and with support that does not put undue stress on the artifact's parts
- Collection monitoring involves regular checks on collections to assess their condition and ensure they are not deteriorating
- Conservation also means the protection and preservation of digital information from archaeological projects
- Ethical aspects of conservation include, where appropriate, consultation with descendant groups and sometimes repatriation of collections to those groups

References cited

Applebaum, B. (2007). *Conservation treatment methodology*. London: Butterworth-Heinemann.

Applebaum, B. (2018). *Preserve, protect and defend: A practical guide to the care of collections*. Scotts Valley: CreateSpace Independent Publishing.

Bleed, P., & Nickel, R. (1989). Optimal management of archeological collections. *The Museum Journal, 32*(1), 26–33.

Bradley, S. (1983). Conservation recording in the British Museum. *The Conservator, 7*, 9–12.

Bradley, S. (1990). *A guide to the storage, exhibition and handling of antiquities, ethnographia and pictorial art* (British Museum Occasional Paper 66). London: The British Museum.

Bradley, S. (1992). Conservation aspects of storage and display. In J. M. Thompson (Ed.), *Manual of curatorship, a guide to museum practice* (2nd ed., pp. 468–473). London: Butterworth-Heinemann.

Bradley, S., & Daniels, V. (1990). Environment. In S. Bradley (Ed.), *A guide to the storage, exhibition and handling of antiquities, ethnographia and pictorial Art* (British Museum Occasional Paper 66) (pp. 1–14). London: The British Museum.

Bradley, S., Uprichard, K., & Munday, V. (1990). General guidelines on the handling of objects. In S. Bradley (Ed.), *A guide to the storage exhibition and handling of antiquities, ethnographia and pictorial art* (pp. 15–17). London: The British Museum.

Childs, S. T., & Benden, D. M. (2017). A checklist for sustainable management of archaeological collections. *Advances in Archaeological Practice, 5*(1), 12–25.

Clarke, R. W., & Bradshaw, S. N. (Eds.). (1982). *Conservation of iron* (Marine monographs and reports 53). London: National Maritime Museum.

Clavir, M., & Moses, J. (2018). *Caring for sacred and culturally sensitive objects*. Ottawa: Canadian Conservation Institute. https://www.canada.ca/en/conservation-institute/services/preventive-conservation/guidelines-collections/caring-sacred-culturally-sensitive-objects.html.

Corfield, M. (1983). Conservation records in the Wiltshire Library and Museum Service. *The Conservator, 7*, 5–8.

Cronyn, J. M. (1990). *The elements of archaeological conservation*. London: Routledge.

Dowman, E. A. (1970). *Conservation in field archaeology*. London: Methuen & Co.

Down, J. L. (2015). The evaluation of selected poly(vinyl acetate) and acrylic adhesives: A final research update. *Studies in Conservation, 60*, 33–54.

Eiteljorg, H. (2004). Archiving digital archaeological records. In S. T. Childs (Ed.), *Our collective responsibility: The ethics and practice of archaeological stewardship* (pp. 67–76). Washington: Society for American Archaeology.

Fink, J. K. (2017). *Chemicals and methods for conservation and restoration: paintings, textiles, fossils, wood, stones, metals, and glass*. New York: Wiley.

Jespersen, K. (1985). Extended storage of waterlogged wood in nature. In *Waterlogged wood: Study and conservation, proceedings of the 2nd ICOM waterlogged wood working group* (pp. 39–54). Grenoble: Centre d'Étude et de Traitement des Bois Gorgés d'Eau.

Keene, S., & Orton, C. (1991). Measuring the condition of museum collections. In G. Lock & J. Moffett (Eds.), *CAA91, Computer applications and quantitative methods in archaeology 1991* (pp. 163–166). Oxford: Tempus Reparatum.

Kemp, J. (2009). Practical ethics v2.0. In A. Richmond & A. Bracker (Eds.), *Conservation: Principles, dilemmas and uncomfortable truths* (pp. 60–72). New York: Routledge.

Kibblewhite, M., Tóth, G., & Hermann, T. (2015). Predicting the preservation of cultural artefacts and buried materials in soil. *Science of the Total Environment, 529*, 249–263.

Kintigh, K. W., & Altschul, J. H. (2010). Sustaining the digital archaeological record. *Heritage Management, 3*(2), 264–274.

Knight, B. (1982). A note on the storage of freshly excavated iron objects. *Conservation News, 17*, 19.

Knoll, M. K., & Carver-Kubik, A. (2019). In-field digital photography and the curation of associated records: Not all prints are created equal. *Advances in Archaeological Practice, 7*(3), 302–310.

Koob, S. (1981). Conservation with acrylic colloidal dispersions. In *Preprints of the AIC 9th annual meeting* (pp. 86–94). Washington: American Institute for Conservation.

Koob, S. (1986). The use of Paraloid B72 as an adhesive: Its application for archaeological ceramics and other materials. *Studies in Conservation, 31*, 7–14.

Leigh, D. (1982). The selection, conservation and storage of archaeological finds. *Museums Journal, 82*, 115–116.

McCawley, J. C. (1977). Waterlogged artifacts: The challenge to conservation. *Journal of the Canadian Conservation Institute, 2*, 17–26.

Meister, N. B. (2019). A guide to the preventive care of archaeological collections. *Advances in Archaeological Practice, 7*(3), 267–273.

Miles, G. (1992). Object handling. In J. M. Thompson (Ed.), *Manual of curatorship, a guide to museum practice* (2nd ed., pp. 455–458). London: Butterworth-Heinemann.

Partington-Omar, A., & White, A. J. (Eds.). (1981). *Archaeological storage*. Society of Museum Archaeologists: Lincoln.

Pearson, C. (1981). The use of polyethylene glycol for the treatment of waterlogged wood: Its past, present and future. In L. H. de Vries-Zuiderbaan (Ed.), *Conservation of waterlogged wood* (Proceedings of the Inernational Syposium on the Conservation of Large Objects of Waterlogged Wood) (pp. 51–56). The Hague: Ministry of Education and Science.

Prendergast, M. E., & Sawchuk, E. (2018). Boots on the ground in Africa's ancient DNA 'revolution': Archaeological perspectives on ethics and best practices. *Antiquity, 92*(363), 803–815.

Pye, E. M. (1992). Conservation and storage: Archaeological material. In J. M. Thompson (Ed.), *Manual of curatorship, a guide to museum practice* (2nd ed., pp. 392–426). London: Butterworth-Heinemann.

Richmond, A., & Bracker, A. (Eds.). (2009). *Conservation: Principles, dilemmas and uncomfortable truths*. New York: Routledge.

Rodgers, B. A. (2004). *The archaeologist's manual for conservation: A guide to non-toxic, minimal intervention artifact stabilization*. New York: Kluwer Academic/Plenum Publishers.

Rose, C. L., & Hawks, C. A. (1995). A preventive conservation approach to the storage of collections. In C. L. Rose, C. A. Hawks, & H. H. Genoways (Eds.), *Storage of natural history collections* (Vol. 1, pp. 1–20). Washington: Society for the Preservation of Natural History Collections.

Shackley, M. S. (1998). *Archaeological obsidian studies: Method and theory*. New York: Plenum Press.

Shearman, F. (1990). Metals. In S. Bradley (Ed.), *A guide to the storage, exhibition and handling of antiquities, ethnographia and pictorial art* (British Museum Occasional Paper 66) (pp. 18–260). London: The British Museum.

Sheehan, B. (2015). Comparing digital archaeological repositories: tDAR versus Open Context. *Behavioral & Social Sciences Librarian, 34*, 173–213.

Skibo, J. M. (1992). *Pottery function. A use-alteration perspective*. New York: Plenum Press.

Stansfield, G. (1985). Pest control: A collections management problem. *Museums Journal, 85*, 97–100.

Sullivan, L. P., & Childs, S. T. (2003). *Curating archaeological collections: From the field to the repository*. Lanham: AltaMira Press.

Sullivan, L. P., & Edwards, A. (Eds.). (2004). *Stewards of the sacred*. Boston: American Association of Museums and Centre for the Study of World Religions, Harvard University.

Tate, J., & Skinner, T. (1992). Storage systems. In J. M. Thompson (Ed.), *Manual of curatorship, a guide to museum practice* (2nd ed., pp. 459–467). London: Butterworth-Heinemann.

Tubb, K. W. (2001). The statues of 'Ain Ghazal': Discovery, recovery and reconstruction. *Archaeology International, 5*, 47–50. http://doi.org/10.5334

UKIC. (1983). *Guidance for conservation practice*. London: United Kingdom Institute for Conservation.

UKIC Archaeology Section. (1982). *Excavated artefacts for publication: U.K. sites. Guidelines 1*. London: United Kingdom Institute for Conservation.

UKIC Archaeology Section. (1983). *Packaging and storage of freshly-excavated artefacts from archaeological sites. Guidelines 2*. London: United Kingdom Institute for Conservation.

UKIC Archaeology Section. (1984). *Environmental standards for the permanent storage of excavated material from archaeological sites. Guidelines 3*. London: United Kingdom Institute for Conservation.

Unger, A., Schnewind, A. P., & Unger, W. (2001). *Conservation of wood artifacts: A handbook*. New York: Springer.

Viñas, S. M. (2009). Minimal intervention revisited. In A. Richmond & A. Bracker (Eds.), *Conservation: Principles, dilemmas and uncomfortable truths* (pp. 47–59). New York: Routledge.

Watkinson, D., & Neal, V. (1998). *First aid for finds* (3rd ed.). London: The British Archaeological Trust and the Archaeology Section of the UKIC.

10 Laboratory Health and Safety

There was a distinct trail of little red spots from the work area to the bathrooms inside the anthropology building

(Whittaker 1994: 79).

An aspect of archaeological laboratory work that is really important yet does not always receive adequate attention in archaeological training is health and safety. It may seem obvious that we need to consider health and safety in, say, chemistry labs, but archaeological labs also present hazards (Poirier and Feder 2001). For example, archaeometric analyses can involve sources of ionizing radiation, preparing sediment samples to examine pollen or phytoliths can involve corrosive acids, and experiments with flint-knapping cause small, sharp flakes to fly in unpredictable directions where they could injure eyes or cut skin. Even though most archaeological laboratory activities have no hazards any worse than what we encounter in our daily lives, it makes sense to pay attention to safety.

Some aspects of laboratory health and safety concern the basic organization and supply of the lab. Generally, creating a safe environment requires that the lab be clean and reasonably tidy, with work counters and seating at ergonomically appropriate heights, with appropriate lighting, with a properly equipped First-Aid kit in plain sight, and without tripping or slipping hazards. Most of the features that lab users should be able to find in a laboratory are listed in Table 10.1, along with others that should be there if the laboratory activities include chemicals, high levels of dust, biohazards (such as dead animals), or flint-knapping (Whittaker 1994: 79–83).

10.1 Ergonomics and Back Strain

A major aspect of ergonomics is the layout and physical specification of lab furniture. Counters and seating should be at the right heights for the kinds of activities anticipated, and there should be no tripping hazards, such as power cables draped across traffic areas.

Moving heavy artifacts, equipment, or trays can easily lead to back injury if not done carefully. Never carry a heavy object across the lab, or between labs, in your hands but instead place it on a lab cart or dolly so that you can wheel it to its destination. When you lift a heavy object, never lift with your back but instead try to keep your back straight and lift with your knees. When you are either removing or replacing a heavy tray that belongs in the upper portion of a storage shelf or cabinet, do not attempt to support its weight above your head. Instead, use a step stool, preferably one of the larger ones with stairs and side-railings, so that you can remove or replace the tray below your own eye level. Some of these step stools even have a rail or shelf where you can rest the tray while you position yourself to lift it into place (Fig. 10.1).

10.2 Hygiene

In any laboratory setting, it is important to avoid letting your handling of artifacts, chemicals, or dirt introduce bacteria or irritants into your mouth or eyes. Keep in mind that most artifacts were once buried in dirt, may not have been cleaned thoroughly, and often accumulate dust while in storage. Do not eat food or consume beverages in a lab, especially one in which there are dirty or dust-creating activities, and be sure to wash your hands prior to eating anything. In some cases, it may be advisable to wear disposal nitrile gloves, which are available in handy tissue-style boxes.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_10

Table 10.1 Elements of a safely equipped archaeological laboratory

Furniture and facilities	
Counters	110 cm (standing), 72 cm (sitting in desk chair) from floor, preferably with height adjustable, with knee spaces for researchers seated on stools. Power outlets should be placed in or above counters to avoid having cables draped across walking areas.
Sinks	Large and deep steel sinks, with silt traps if the laboratory activities include washing artifacts or wet-sieving sediment
Stools	~60 cm high for use at counters
Windows	For natural light and reduction of eye strain
Fume hood	To evacuate smells and hazardous gases or aerosols from organic remains, chemicals, adhesives, kilns or lab ovens
Eye-wash station	Essential in labs that employ chemicals that could be damaging to eyes
Step stool	To facilitate retrieval of trays or artifacts on high levels of storage cabinetry
Chemical Storage Cabinet(s)	Fume-proof, fire-proof, explosion-proof. Note that different kinds of chemicals (acids, bases) need to be in different cabinets
Lab cart	For moving heavy trays or artifacts without risk to lab workers or artifacts
Dolly	For moving heavy boxes, furniture or lab equipment
Equipment and supplies	
First-Aid Kit	Should contain bandages and anticipate other possible injuries in lab
Materials Safety Manual	Binder with documentation of any chemicals used in lab, instructions for clean-up, emergency phone numbers
Spill Kit	For cleaning up the kinds of spills anticipated in lab
Lab coats	For protecting researchers' clothing from dirty or hazardous artifacts or materials – should be washed regularly
Nitrile gloves	For handling some kinds of artifacts and all chemicals. Available in convenient boxes
Work gloves	For handling sharp or hot objects and for flint-knapping
Goggles	To protect eyes from chemicals, dirt (during screening) or flakes (during flint-knapping)
Face masks	To protect lungs from dust during screening of sediments, or eyes from dust, chemical splashes, or flying debris
Broom and dust-pan	Have different (labelled) ones for materials that should not be mixed together (e.g., chemical spills, foreign soils)
Waste disposal	Include dedicated disposal containers for acids and bases, if needed, as well as for biohazards (e.g., waste from preparing skeletons), foreign soils, and other things that should not go into normal landfill

10.3 Eye Protection

Safety goggles are the most basic requirement to protect eyes during such laboratory activities as flint-knapping. They are also useful during very dusty activities, such as dry-screening sediments.

Eye strain is one of the most common injuries in archaeological laboratories. It is particularly serious when lab activities include a good deal of microscopic observation, as in archaeobotany (Fig. 10.2). To protect against eye strain, not only does the lab require excellent lighting, preferably including natural light from a window, but it is important to coach all lab workers about the importance of varying their focal distance regularly and taking breaks between sessions of microscope work, rather than staring too long through a microscope or looking closely at hundreds of artifacts without a break. Periodically, each observer should look up and focus on something farther away, on the other side of the room or out the window.

10.4 Cuts and Abrasions

Flint-knapping is the archaeological activity most likely to cause cuts and those participating in that activity should *always* wear both goggles and work gloves. Other archaeological work that involves using sharp instruments or grinding equipment, such as rock saws, polishing wheels, or scalpels can also lead to injuries.

The debris from flint-knapping typically contains small, sharp flakes, but it is safe for disposal in ordinary landfill, although you should avoid inhaling dust from the debris while sweeping it up. Knapping debris, crushed in a mortar, can also be excellent temper for pottery-making.

10.5 Respiratory Hazards

Some archaeological activities raise a lot of dust, in which case the most basic protection is a dust mask. Dust masks should have two straps, not one, so the wearer can adjust it properly across the mouth and nose.

However, dust masks are not effective protection from chemical fumes, which instead require use of a fume hood, an appropriately rated face mask, or both.

Fume hoods are not all the same. Some vent through a chimney; others recirculate air through a filter. In both cases, the fume hood needs to be rated for the kinds of fumes that we can reasonably expect our lab work to generate. This can range from acids, through acetone, to the gases generated from preparing skeletons in a zooarchaeological lab.

Fig. 10.1 Laboratory step-stool for accessing high shelving

10.6 Radiation Hazards

Archaeologists sometimes use X-ray fluorescence (XRF and pXRF), X-ray diffraction (XRD), neutron activation (INAA), and other analytical methods that depend on ionizing radiation or radioactivity to determine the chemical, elemental or isotopic composition of ancient materials. Often this takes place in a specialized laboratory, where there are safeguards, such as lead shielding and use of radiation monitors (dosimeters), but the recent availability of portable XRF has brought some of these methods, along with the need to ensure their safety, into regular archaeological laboratories. In many jurisdictions, even using portable XRF requires training and certification.

10.7 Workplace Hazardous Materials

Although most archaeological laboratories either use no chemicals or use only a small number of fairly innocuous ones, it is more common than some people realize for archaeological labs to contain potentially dangerous materials. For example, many of the adhesives that archaeologists use are acetates dissolved in acetone; acetone is a carcinogenic substance that requires careful handling and disposal. The desiccant, silica gel, is also hazardous if inhaled and is an irritant to skin or digestive tracts. Meanwhile, some archaeological lab activities, and especially those involving "digestion" of sediments for pollen analysis or geoarchaeology, involve highly corrosive acids.

Zooarchaeological laboratories can sometimes have biohazards, at least temporarily, during preparation of

Fig. 10.2 Using microscopes at ergonomic heights with plenty of illumination from a large window

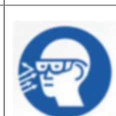
skeletons from carcasses (see Chap. 15). Here, the biohazard consists of rotting or putrid flesh, often mixed with various chemicals. By contrast, prepared bones, archaeological bones, or fresh animal parts used in archery, butchering, cooking, or tool-making experiments are no more hazardous than the meat in your kitchen or the contents of a restaurant's trash. Use careful hygiene while handling such materials, just as you would in your kitchen, but disposal of such material requires no special protocols.

Whenever you have dangerous substances, whether chemical or biohazards, in your lab, even temporarily, you should post the hazard descriptions from the Globally Harmonized System of Classification and Labelling of Chemicals (GHS 2017) or its equivalent in your country (e.g., OSHA, WHMIS), ensure that everyone working with these materials reads them, and put spill kits and instructions for using them in accessible locations in the work area (OSHA 2014; WHMIS 2015a, d). There should also be eye-wash stations in places where any of these more hazardous chemicals are used.

10.8 Hazard Warning Labels

Table 10.2 displays some of the symbols that are most relevant to the kinds of hazards likely to be found in archaeological laboratories. The standard icons vary by jurisdiction, but most have recently adopted a version of the Globally Harmonized System for Classifying and Labelling Chemicals (GHS 2017). Hazard warnings are standardized by OSHA (2014) in the United States, the Classification, Labelling and Packaging (CLP) regulation in the European Union (ECHA 2015), the Workplace Hazardous Materials Information System (WHMIS 2015b, 2015c) in Canada, and the Hazardous Chemical Information System in Australia (HCIS; Safe Work Australia n.d.).

Table 10.2 A selection of hazard labels used in the United States, Europe or Canada. Many others are similar

	Hazard or amelioration		Hazard or amelioration
	Flammable material		Serious health hazard (substances that are not acutely toxic but can have long-term health impacts)
	Gases under pressure		Biohazard (e.g., refuse from processing animal carcasses)
	Corrosive (e.g., HCl acid)		Radiation hazard (ionizing radiation, e.g., pXRF)
	Oxidizer		Environmental hazard (e.g., foreign soils that should not be dumped without prior sterilization)
	Explosive		Wear protective gloves (e.g., flint-knapping, use of lab ovens)
	Skin, eye or respiratory irritant		Risk of eye injury, wear safety glasses (e.g., flint-knapping)
	Acute toxicity		

10.9 Summary

- All laboratories, including archaeological ones, pose hazards, so ensuring health and safety is very important.
- General rules are that labs should always have First-Aid kits, be furnished ergonomically, and have hazard labels that draw attention to any specific risks, such as chemicals, dust, or decomposing animal carcasses.
- Those who work in archaeological laboratories should be familiar with any hazards that are involved in the laboratory work and be trained to minimize risk.

References Cited

ECHA. (2015). *Understanding the Classification, Labelling and Packaging (CLP) regulation*. European Chemicals Agency. https://echa.europa.eu/regulations/clp/understanding-clp. Accessed 20 Dec 2018.

GHS. (2017). *Globally harmonized system of classification and labelling of chemicals (GHS)* (17th revised ed). New York: United Nations Economic Commission for Europe. www.unece.org/trans/danger/publi/ghs/ghs_rev07/07files_e.html#c61353. Accessed 20 Dec 2018.

OSHA. (2014). *Hazard communication: Small entity compliance guide for employers that use hazardous chemicals* (OSHA 3695-03 2014). Washington, DC: Occupational Safety and Health Administration, US Department of Labor.

Poirier, D. A., & Feder, K. L. (2001). *Dangerous places: Health, safety, and archaeology*. Westport: Bergin & Garvey.

Safe Work Australia. (n.d.). *Hazardous Chemical Information System (HCIS)*. www.hcis.safeworkaustralia.gov.au. Accessed 20 Dec 2018.

Whittaker, J. C. (1994). *Flintknapping: Making and understanding stone tools*. Austin: University of Texas Press.

WHMIS. (2015a). *WHMIS 2015 – Guide for employers and workers*. Ottawa: Canadian Centre for Occupational Health and Safety.

WHMIS. (2015b). *WHMIS 2015 – Labels*. Ottawa: Canadian Centre for Occupational Health and Safety.

WHMIS. (2015c). *WHMIS 2015 – Pictograms*. Ottawa: Canadian Centre for Occupational Health and Safety.

WHMIS. (2015d). *The safety data sheet: A guide to first-aid recommendations*. Ottawa: Canadian Centre for Occupational Health and Safety.

Part II

Introduction to Part II

Keywords
Interdisciplinary archaeology; Collaboration; *Chaîne opératoire*Taphonomy

This part moves on to specific categories of artifacts, animal and plant remains, and kinds of analyses in archaeology. While it is necessary to treat these separately, it is important to realize that the best archaeology integrates data from multiple sources and is inherently collaborative (Stutz 2018). For example, research on food preparation and consumption inevitably must take into account multiple strands of evidence—zooarchaeological, archaeobotanical and artifactual—in order to build a coherent argument (Graff 2018). Increasingly, archaeology is also collaborative in the sense of cooperation with indigenous and descendant communities and de-colonizing the study and representation of the past (La Salle and Hutchings 2016; Nicholas et al. 2011; Ray 2009; Silliman 2008).

All the topics in Part I and several additional themes are relevant to most or all of the following chapters. You will note the use of the *chaîne opératoire* (defined pp. 159–160), as one of the paradigms guiding analysis of lithics, pottery, metal, bone and shell artifacts, as well as plant remains. Taphonomy, the study of the processes that affect the deposition, preservation, distribution, recovery, and interpretation of remains in archaeological deposits, is a topic that has a particularly important role in the chapters on zooarchaeology and archaeobotany, but emerges in the analysis of all kinds of archaeological remains, as already noted in the chapters on research design and quantification.

Many of the chapters in this section focus quite a lot on terminology for the description and identification of artifacts, animal and plant remains, including their “anatomy” and major categories. These provide the “information language” for these materials. Here, you should recall from Chap. 3 that many of these terms are related to systematics or describe attributes that we can use to classify or group things. Still other terms set out orientation and segmentation rules as discussed in Chap. 4, using terms such as “transverse,” “distal” and “lateral,” that are relevant to most of our materials.

Many of the chapters also discuss briefly, or illustrate with case studies, some kinds of research question that archaeologists have pursued on the basis of these kinds of material evidence. This is, of course, not exhaustive, but my goal was to give a taste of some of the major themes and theoretical paradigms.

Finally, and throughout the book, issues of data quality and reliability and their role in the persuasiveness of archaeological arguments continue to be a focus.

References Cited

Graff, S. R. (2018). Archaeological studies of cooking and food preparation. *Journal of Archaeological Research, 26*(3), 305–351.

La Salle, M. J., & Hutchings, R. (2016). What makes us squirm: A critical assessment of community-oriented archaeology. *Canadian Journal of Archaeology, 40*(1), 164–180.

Nicholas, G. P., Roberts, A., Schaepe, D. M., Watkins, J., Leader-Elliot, L., & Rowley, S. (2011). A consideration of theory, principles and practice in collaborative archaeology.

Archaeological Review from Cambridge, 26(2), 11–30.

Ray, C. (2009). Emerging consensus and concerns in collaborative archaeological research. *Collaborative Anthropologies, 2*, 1–8.

Silliman, S. W. (2008). Collaborative indigenous archaeology: Troweling at the edges, eyeing the centre. In S. W. Silliman (Ed.), *Collaborating at the trowel's edge: Teaching and learning in Indigenous archaeology* (pp. 1–21). Tucson: University of Arizona Press.

Stutz, L. N. (2018). A future for archaeology: In defense of an intellectually engaged, collaborative and confident archaeology. *Norwegian Archaeological Review, 51*(1–2), 48–56.

11 Stone Artifacts

The chief significance of stone tools, in themselves and in their context on a living floor, is for what they can tell us about the way of life of the people who made them

(Clark 1970: 136).

Among our best-preserved and most ubiquitous evidence for most of the human career are stone tools and the debris from making them. It is not surprising that archaeologists have spent so much time analyzing lithics (Andrefsky 2005; Goodale and Andrefsky 2015; Odell 2012; Shea 2013; Whittaker 1994).

Broadly speaking, lithics consist of two major categories: flaked or chipped stone and ground stone, although these categories are artificial and not entirely distinct (Adams 2002; Rosenberg et al. 2016; Rowan and Ebeling 2008). When archaeologists refer to "lithics," they are typically referring to tools that display scars from flake removals. When they refer to "ground-stone tools," they refer to tools that were not necessarily manufactured by grinding or used for grinding, but ones that, at least in finished form, display few or no flake scars from manufacture, at least on their use surfaces, although they may display scars from use.

There are many ways to approach research on stone tools and their technology, including a long-standing typological one that focuses on the form and retouch of finished tools, and one borrowed from engineering called decision theory (e.g., Bleed 1986; Kingery 2001; Kukan 1978). Many lithic researchers take a primarily technological approach, often emphasizing the sequence of flake removals or other activities in the process of making artifacts (e.g., Schiffer 1976). This reduction-sequence approach is still common, especially in North American archaeology, sometimes characterized as a "behavioral chain" that includes the entire process from raw material to eventual discard (Skibo and Schiffer 2008: 9–10). A related approach places its focus on technological organization (Nelson 1991). This involves examining the "dynamics" – the plans and strategies – of technological behavior in the context of particular resource, social and economic environments that lead to particular tool designs and staging and distribution of manufacturing activities.

11.1 *La Chaîne Opératoire*

An approach that originated among French archaeologists but has been increasingly adopted in North America and elsewhere is superficially similar to behavioral-chain or technological-organization approaches. This is the ***chaîne opératoire***, sometimes anglicized as "operational chain" (Leroi-Gourhan 1964). There is no "official" definition of this approach (Audouze et al. 2017; Audouze and Karlin 2017), Leroi-Gourhan having treated it as a concept, while others have treated it more as a methodology (Maget 1962; Tixier 2012). Most lithic applications of *chaîne opératoire* have focused on the manufacturing process, while others note the importance of all the processes involved in people's use of materials, from discovery, selection, acquisition and processing of raw materials, through manufacture, use, and reuse of artifacts, to recycling and eventual discard. Although it may seem similar to the reduction-sequence approach in that it most of its applications focus on manufacturing sequences (Bleed 2001; Shott 2003), the *chaîne opératoire* approach additionally refers to the strategies that people use, their technical knowledge and "know-how," the decisions they make at each step, and the gestures they have learned through immersion in a culture or community of practice, or

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47992-3_11) contains supplementary material, which is available to authorized users.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_11

by apprenticeship to others (Andrefsky 2005; Chazan 2009; Pelegrin 1993). In that respect, it bears some similarities to the technological-organization approach, which also considers social environment and strategies.

The *chaîne opératoire* provides a framework for understanding variability in lithic and other assemblages by helping us "read" the signs left by tool makers' and tool users' decisions (Audouze et al. 2017; Chazan 1997; Inizan et al. 1992: 12–13; Lemonnier 1986; Schlanger 2004; Sellet 1993; Soressi and Geneste 2011), but there is a lot more to it than a sequence and, in some technologies at least, elements of the sequence can occur simultaneously (Chazan 2009). Notably, there is an inherent assumption of intentionality and goal-orientation in most applications of *chaînes opératoires* (e.g., Dibble 1995: 304); yet it is also true, even in a *chaîne-opératoire* framework, that operations like resharpening and re-use can cause final products to differ remarkably from what a flint-knapper may originally have envisioned. In the case of lithic technology, users of the *chaîne opératoire* consider the characteristics of tools, manufacturing debris, and other debris as the result of physical actions and people's motor abilities, skills (*savoir faire* or know-how), conceptual knowledge (*connaissance*) and experience, as expressed in ability and performance. Consequently, *chaînes opératoires* also have a relationship to cognitive archaeology (Mahaney 2014).

While it is common for researchers to study the *chaînes opératoires* of a particular aspect of technology, such as lithics or even a particular tool type, some practitioners of this approach emphasize that flint-knapping and other forms of stone reduction are subsystems of a larger technological system (Lemonnier 1986: 148), which of course includes technologies discussed in other chapters in this book.

11.2 Lithic Raw Material

One of the principal limitations on the manufacture of stone tools is raw material. The makers of tools appreciated that different types of rock varied in ease of flaking or pecking, sharpness and longevity of edge, or ability to withstand grinding or pounding, aesthetic appearance, and other desirable qualities. One clue to this is that tools that appear to have had different functions often are made of different materials, with basalt and quartzite, for example, used for heavy chopping tools or for grinding tools, obsidian for making very sharp knives or finely retouched tools, and flint or chert for more general cutting and scraping tools. Tool makers sometimes favored raw materials they could not obtain locally, requiring them to acquire them from hundreds of kilometers away.

The most commonly used material for chipped-stone tools is **flint** or **chert**. This is a micro-crystalline (usually 2-50μ) silicate (SiO_2) rock, formed as nodules or layers in limestone or chalk, that behaves in many ways like a super-cooled fluid (Hodges 1964: 99), similar to glass. The most important feature of such materials, from the perspective of tool production, is that they break with conchoidal fracture, as explained below. Flint forms over millions of years in ocean sediments. Broadly defined, any sedimentary rocks composed mainly of microcrystalline quartz can be considered chert, including flint, chalcedony, agate, jasper, and hornstone (Luedtke 1978).

Where knappers require an extremely sharp edge or great control over fracture mechanics, they prefer **obsidian**, if it is available. It is a volcanic (igneous) glass, an amorphous silicate that, like all glass, is a super-cooled fluid. This means that, because obsidian cools so rapidly in volcanic flows, it lacks a crystal structure. Obsidian flakes have extremely sharp edges, but they are also quite brittle.

Quartzite is a hard, metamorphic rock that does not flake as easily as chert, but is still useful, especially as orthoquartizite, for some heavy tools, such as choppers and hoes, as well as for making grinding stones.

Quartz is a crystalline silicate (SiO_2) that is hard and holds a sharp edge but often has flaws in its crystal structure that make it very hard to flake predictably. Consequently, it is usually only used as material for flaked-stone tools where cherts and obsidian are unavailable, and generally for tools that are relatively small in size. It can also be ground and polished to make decorative or luxury items.

Basalt, rhyolite, andesite, dolerite and gabbro are igneous rocks composed mainly of plagioclase and pyroxene minerals (Andrefsky 2005: 46–48), rhyolite also being high in silica, which improves its fracture characteristics. These and most other igneous rocks besides obsidian do not flake as well as chert but are excellent materials for ground-stone tools. Most archaeologists do not explicitly distinguish between basalt, which is more fine-grained, and gabbro, which is coarse, and sometimes even include pumice, which is a porous glass formed during explosive volcanic eruptions, among basalts. Despite its flaking difficulties, knappers sometimes used basalt, and especially rhyolite, to make bifaces, but it is much more commonly the material for ground-stone tools.

Although hard, dolomitic **limestone** is not as suitable for flaked-stone tools, it is excellent for making building stones and stone vessels and mortars. Makers of these things can use flaking to rough out a shape and then pecking and grinding to finish it.

A number of other sedimentary and metamorphic rocks are also excellent material for ground-stone tools, decorative items, or vessels, including alabaster, calcite, chlorite, gneiss, marble, serpentinite, schist, and steatite (Phillips and Simpson 2018).

11.2.1 Sourcing the Raw Materials

Archaeologists have been very successful at sourcing obsidian tool-stone, while their success at sourcing other stone materials has been variable. To some extent, archaeologists, no doubt like prehistoric toolmakers, can distinguish different materials by their color, banding, translucence, inclusions, grain, and other characteristics. However, archaeologists have also turned to analytical chemistry, and sometimes microfossils, to distinguish sources more reliably (Braswell et al. 2000; Glascock 2002; Luedtke 1979; Sánchez de la Torre et al. 2019).

Trace elements (elements in the parts-per-million or ppm range) in obsidian have long helped archaeologists distinguish the sources of archaeological obsidians. A classic early example was Dixon et al.'s (1968) analysis of obsidians originating in what is now Turkey to make inferences about obsidian trade in the ancient Near East. Obsidian sourcing has since then been important in Mesoamerica (e.g., Carballo et al. 2007), northwestern North America (e.g., Reimer 2015), Japan (e.g., Kuzmin et al. 2013), the South Pacific (e.g., Golitko et al. 2010), northeast Africa (e.g., Blegen 2017; Shackley and Sahle 2017), and elsewhere. Obsidian sourcing can depend simply on scatterplots of the ratios between key elements as measured with X-ray fluorescence, neutron activation or laser-ablation ICP-MS, or on multivariate analyses of the results of the elemental concentrations (e.g., Fig. 11.1). As source identifications are inductive, and sources may not be uniform in their chemistry (e.g., Hughes 1994), the probability that identifications are correct depends in part on how well the known sources have been sampled, and how many potential sources remain unknown.

Geochemistry has also helped source the materials of other artifacts made from igneous rocks, such as basalts (e.g., Bostwick and Burton 1993; Rutter et al. 2003; Weisler et al. 2016). To date, it has not been as successful at identifying sources of chert or flint, often because there is more variability within than between flint sources, although there has been recent progress (Brandl et al. 2018; Moreau et al. 2016; Speer 2014).

11.3 Chipped-Stone Tools

11.3.1 Fracture Mechanics and the Manufacture of Flaked-Stone Tools

Lithic manufacture is fundamentally a reductive technology. That means that the manufacturer shapes it by removing material from a core or flake of stone (Cotterell and Kamminga 1990). For chipped-stone tools, the knapper does this by striking a core or flake at a particular angle and location, or applying sudden pressure at that location, to cause **conchoidal fracture** (Fig. 11.2). This involves creating a shock wave that radiates through the material from the point of impact or pressure in a widening or conical wave front, or "Hertzian cone," the same feature that results from a bullet striking plate

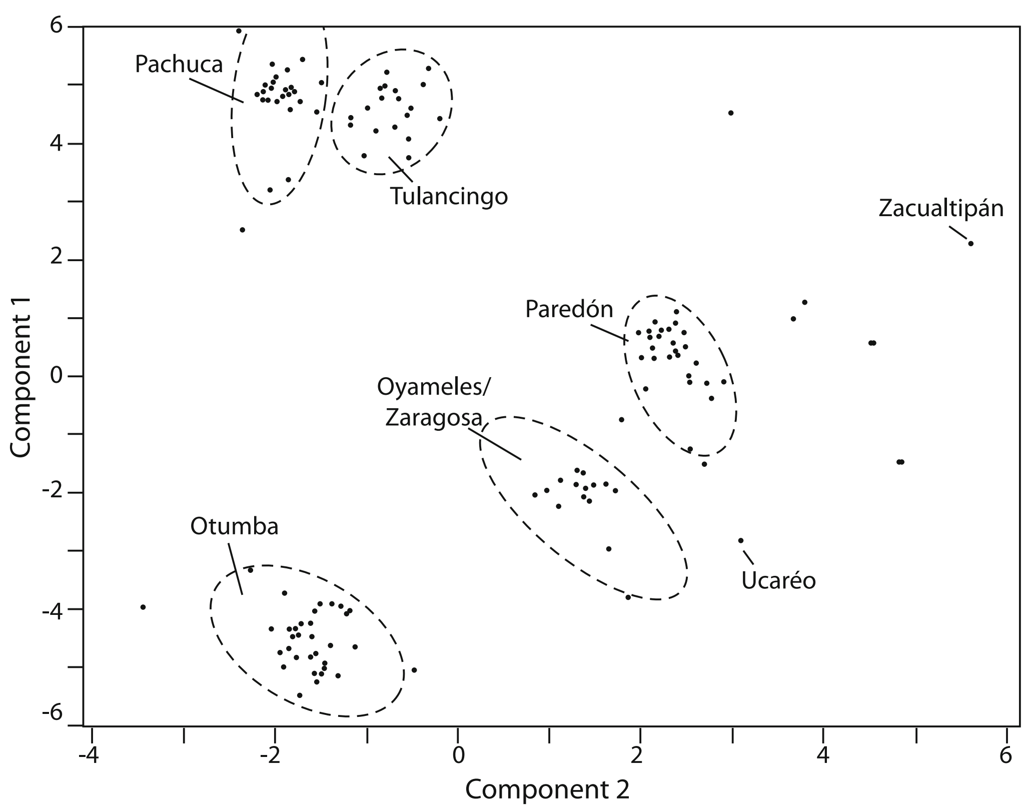


Fig. 11.1 Plot of the first two Principal Components of archaeological samples and quarry pieces analyzed by ICP-MS, a type of mass spectrometry, with ellipses for the 95% confidence intervals for group membership. (After Carballo et al. 2007: 36)

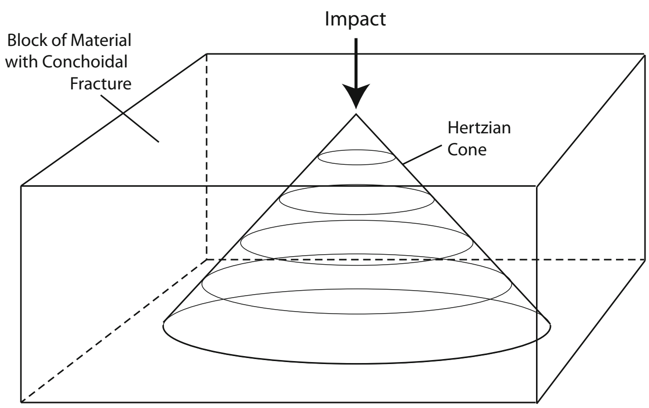


Fig. 11.2 Conchoidal fracture, resulting in a Hertzian cone that expands from the point of impact by a hard object on a block of glass or flint

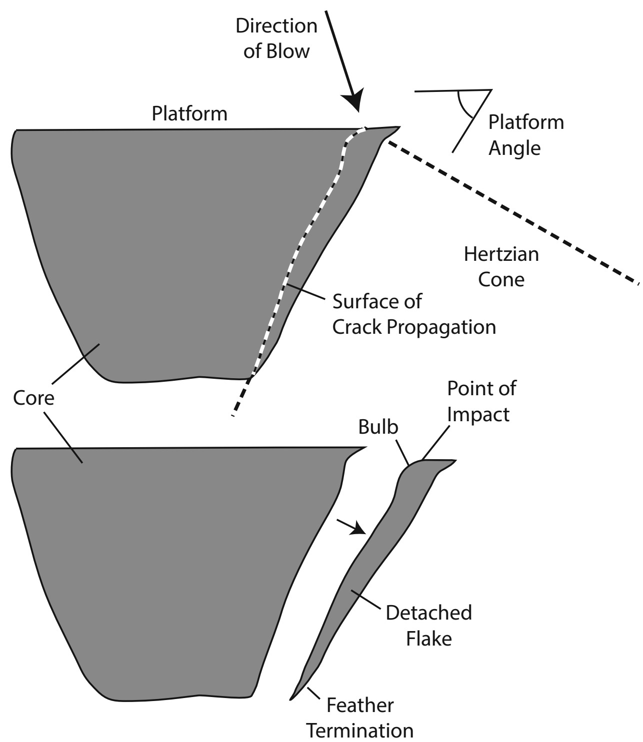


Fig. 11.3 Removal of a flake from the side of a core by conchoidal fracture

glass (Miyamoto and Murakami 2000; Speth 1972). In flint, the angle between the sides of the cone is usually between 120° and 160° (Hodges 1964: 99), and knappers exploited this predictability to make thin flakes by directing a blow near an edge with an angle less than 90°. This truncated the cone so that the resulting flake had a surface nearly parallel to a surface of the core (Fig. 11.3). Whether or not a raw material was suitable for conchoidal fracture typically determined whether toolmakers used it for chipped or only ground-stone tools.

Knappers sometimes shaped or prepared cores in such a way that they could make the removal of flakes or blades (elongated flakes) more predictable. To make a **pyramidal core** suitable for removing many blades, for example, they could take a roughly egg-shaped cobble and knock off one end by a sideways blow with a heavy hammer (Fig. 11.4). This left a flat surface, called a platform, onto which the knapper would direct subsequent blows where the angle between the platform and the sides of the cobble is close to 75°. Striking the platform near the edge at an angle of about 75° in the opposite direction causes the conical wave front to pass through the stone at an angle that roughly parallels the face of the cobble. This causes a relatively thin flake to detach. Continuing this process around the perimeter of the cobble results in a sequence of relatively long, vertical scars so that the cobble becomes a core that is roughly like an upside-down pyramid or faceted cone, although it is usually necessary to remove some core-maintenance flakes to maintain the desired shape and platform angle. For a while, the raised ridges between flake removals on the core help to guide the flake removals and make them more predictable. As the knapper continues to remove flakes, however, the angle between the platform and the sides of the core eventually becomes too great, or the core gets too small, and the knapper rejuvenates it by knocking off a large flake to create a new platform (Fig. 11.4c), recycles it to make a core tool, or discards the "exhausted" core in favor of a new one.

Sometimes knappers used more than one platform on the same core, sometimes removing flakes from one, then another, either alternately or as circumstances suggested.

Archaeologists distinguish several techniques of core reduction, the removal of **debitage** (flakes and other products) from cores.

Bipolar reduction is a simple but unpredictable way to produce many flakes, chips, and chunks with little or no control over their size or shape. It involves placing a core on a large stone used as an anvil and striking the core very hard from above with a large hammer to shatter it. One would then select any usable flakes from those that result, leaving considerable waste. Bipolar technique is also useful for reducing very small cores that would be difficult to flake otherwise.

Hard-hammer percussion involves striking the core near the edge of the platform with a stone hammer, such as a rounded pebble, as in the discussion of conchoidal fracture above (Fig. 11.5).

Soft-hammer percussion involves striking the core in much the same way, but with a hammer made of antler, bone, hard wood, or some other material softer than stone. This kind of hammer is called a billet, baton, or percussor. It is particularly useful for biface manufacture.

In **indirect percussion**, the hammer does not strike the platform at all, but instead the end of a punch whose distal end rests on the platform's edge. This technique allows a

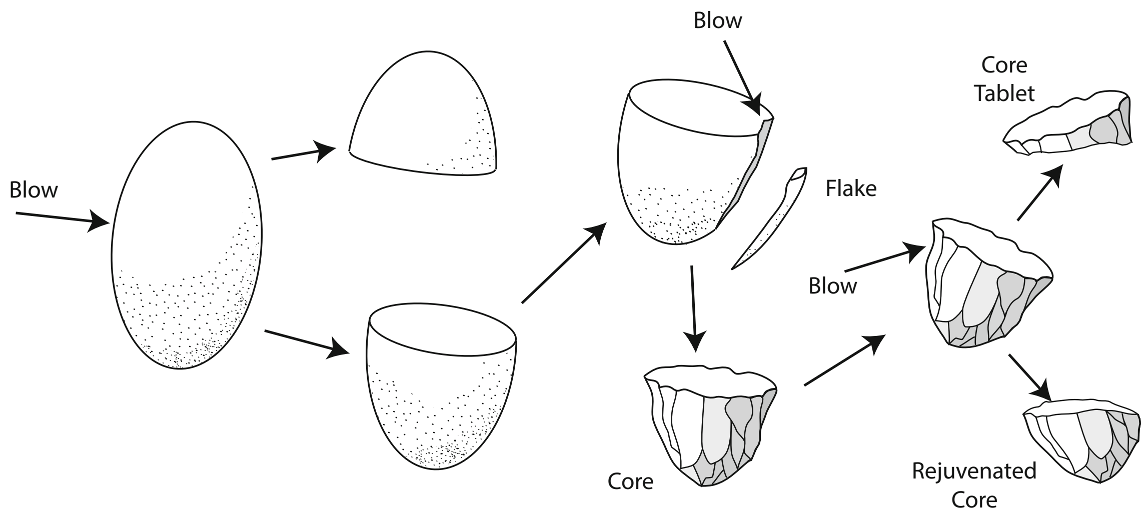


Fig. 11.4 Making a pyramidal core from an egg-shaped cobble (**a**, **b**) and rejuvenating the core (**c**)

Fig. 11.5 A student attempting to learn hard-hammer reduction

knapper to control the location and angle of force much more accurately than the previous techniques, thus improving the predictability of the resulting flake or blade. Punches can be made of antler, but antler wears out quickly. Modern flintknappers often use copper-tipped punches, which of course would not have been available to most prehistoric knappers. Indirect percussion can produce very long blades.

Pressure flaking involves no percussion, but instead sudden pressure with a somewhat pointed tool. Most commonly, this consists of pressing a flaker perpendicular to the edge of a flake to snap off small flakes to **retouch** a tool's edges. This fine-tunes a tool's shape, can be used to sharpen or blunt certain edges, and even to thin the flake. In more dramatic uses of pressure flaking, knappers used their upper-body mass or a large lever to direct sudden pressure along a large flaker placed on the platform edge of a blade core. This technique takes considerable skill, but can result in long, very regular blades.

Cores can just be the source of debitage or can be shaped into tools themselves. Tools made from cores are called **core tools**, often taking the form of **bifaces** (core tools flaked on two sides).

11.3.2 The Anatomy of Chipped Stone

Lithic analysts broadly share some terms for the products of flintknapping and for features on those products (Ballin 2000; Hranicky 2013; Inizan et al. 1992; Whittaker 1994).

The main products of most kinds of flintknapping are cores or core tools and **flakes** removed from cores. Flakes that a knapper has selected to make into flake tools are called **blanks**. **Blades** are flakes that are at least twice as long as they are wide. In France and among users of the *chaîne opératoire*, **debitage** is the term for all material removed from a core, including debris from shaping a prepared core, flakes that could be used as tool blanks, and unusable flakes, chips and chunks called debris (Inizan et al. 1992: 84). However, other lithic analysts use the term "debitage" differently, to refer to waste products of core reduction, including discarded cores. To avoid confusion, this book will refer to debris and discarded cores as waste, and reserve "debitage" for its French usage. Parts of broken flakes and cores are called **fragments**, and sections of intentionally snapped blades are called **segments**.

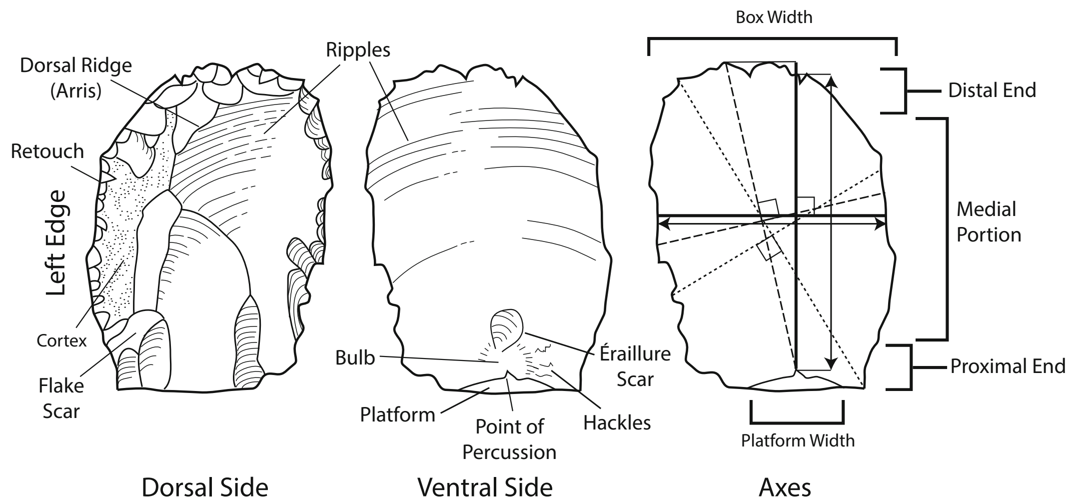


Fig. 11.6 Anatomy of a flake or blade, with some of its main measurements. In the ventral view at right, the heavy lines represent the "box length" and width at half box length, while box width is shown above. Box length and width define a rectangle that encloses the flake when aligned to the platform. Axes with arrows represent the axial length and axial width at half-length, axes with long dashes represent long axis and width at half long axis, and lightly dashed lines represent maximum length and width at half maximum (which in this case is also equal to maximum width)

Flakes and cores exhibit some common features. Among these are **cortex**, the weathered surface that originally covered the stone used for a core, the **platform** (on flakes, only a small remnant of the platform), and **scars** that result from previous flake removals (Fig. 11.6).

Complete flakes and blades show several distinctive features that help us distinguish them from naturally broken stone or glass. At the **proximal end**, we find a small portion of the striking platform. Next to this, on the **ventral surface** (the face that was detached from the core), we usually find a small lip and the **bulb of percussion**, a raised bump or distorted portion of the Hertzian cone that results from compaction under the force of striking, especially by the hard-hammer technique. Sometimes several lines on the bulb, called **radial fissures** (or hackle marks), radiate from the point of percussion on the platform, while hard-hammer percussion sometimes also leaves an *éraillure* scar on the bulb. As long as the material is fine-grained, you will probably see **ripples** (Wallner lines) on the ventral surface that extend away from the point of percussion like waves in a still pond after someone throws in a pebble. These may continue to the **distal end**. The side of the flake that was on the outer part of the core when someone struck off the flake is called the **dorsal surface**. It may show cortex over some or all of it if the flake was one of the earlier ones removed from the core, in which case it is a **cortical flake** or **primary flake**. All but the earliest flakes will show signs of previous flake removals, called flake scars. The sharp ridge that marks the border between flake scars is called an **arris** or **dorsal ridge**.

Although some naturally broken stones may show bulbs of percussion, ripples, or radial fissures, any assemblage of flakes that shows these features consistently is almost certainly the product of knapping by humans or hominins. Most naturally shattered flint exhibits many right angles and flat surfaces that follow planes in the material's crystalline structure, rather than the wave-front of conchoidal fracture.

11.3.3 Stone Tool Typology

From the time that nineteenth-century antiquarians routinely recognized prehistoric stone tools as tools, rather than some natural phenomenon, they began to classify them (see Chap. 3). Typically, they employed either formal typologies, based on the artifacts' shapes in plan view, or they adopted functional typologies, with categories based on the presumed uses of the tools. This typological heritage is still with us, as we continue to use formal terms like "biface," "triangle," or "trapeze" for certain types of stone tools, but also functional terms, such as "axe," "projectile point," and "quern." Many typologies used today actually have a mixture of formal and functional categories (e.g., Bordes 1961; Hranicky 2013; Shea 2013), but lithic analysts also categorize tools in terms of their technology and chronology and study distinct attributes of tools that they think will help them infer function or other aspects of the tools, such as their reduction sequences, *chaînes opératoires*, or extent of resharpening.

11.3.4 Attributes of Chipped-Stone Products

Among the infinity of potentially measurable attributes on any lithic artifact, archaeologists routinely measure some that they believe are relevant to understanding how tools were made, what they were designed to do, and to what period, complex or culture an assemblage might belong. They also measure ones that help them identify sources of raw materials, to discover how tools were used, to infer whether a site was a temporary camp or a long-term settlement, and to address may other research questions. Some attributes occur on all kinds of flakes or blades, others only on certain kinds of tools, and others are specific to waste products (Andrefsky 2005; Burton 1980; Dibble 1985).

11.3.4.1 Attributes Common to Most Lithic Artifacts

The most obvious of these are characteristics of the raw material, such as rock type, color and texture. This can be important, not only for identifying sources, but for evidence that tool makers selected particular materials for different types of tools, or for different reduction strategies. However, archaeologists also make observations on features that are relevant either to preparation of the material, such as heat treatment (Domanski and Webb 2007), or to post-depositional alterations, such as evidence of burning, trampling, or erosion (Table 11.1).

11.3.4.2 Attributes and Classes of Cores

Archaeologists mainly find exhausted cores, discarded once they approached or reached the stage of not being useful. Among the attributes we can observe on such cores are ones

Table 11.1 Examples of lithic attributes along with commonly used scales of measurement and typical purpose

Attribute	Scale	Purpose
All lithics		
Raw material	Nominal	Sourcing, selectivity, economy
Color	Nominal	Identification of heat treatment
Surface lustre	Ordinal	Identification of heat treatment
Pot-lid fractures	Dichotomous	Identification of heat treatment
Presence of cortex	Nominal	Indicates stage of reduction
Extent of cortex	Ordinal/ratio	Indicates stage of reduction
Condition	Nominal or ordinal	Identifying damage from post-depositional factors
Cores		
Number of platforms	Ratio	Identification of reduction strategy
Core shape	Nominal	Identification of reduction strategy
Core type	Nominal	Indicates intended products
Flakes and blades		
Axial length	Ratio	Size and shape, function
Maximum axial width	Ratio	Size and shape, function
Width at half length	Ratio	Size and shape, function
Maximum thickness	Ratio	Indicator of reduction technique
Thickness at midpoint	Ratio	Indicator of reduction technique
Prominence of bulb	Ordinal	Indicator of reduction technique
Termination	Nominal	Identifies knapping errors
Platform type	Nominal	Related to hammer type and reduction technique
Exterior platform angle	Ratio	Related to reduction technique
Platform width	Ratio	Related to reduction technique
Platform depth	Ratio	Related to original flake size
Platform preparation	Nominal	Related to reduction technique
Scar orientations	Nominal	Distinguishes single- from opposed-platform and multidirectional reduction
Retouched tools		
Tool type	Nominal	Related to use or function
Retouch type	Nominal	Related to hafting and function
Locations of retouch	Ordinal	Related to hafting and use
Invasiveness of retouch	Ordinal/ratio	Related to use, curation, style
Steepness of retouch	Ratio	Related to hafting, function, and degree of curation/resharpening
Length of retouch	Ratio	Related to function
Presence of Polish	Dichotomous	Indicative of use
Haft width	Ratio	Relevant to hafting design
Haft length	Ratio	Relevant to hafting design
Haft type	Nominal	Relevant to hafting design
Height of notch	Ratio	Relevant to hafting design

See Andrefsky (2005), Shea (2013: 334–345), and Whittaker (1994)

Note that many of the ratio-scale measures are also used to create indirect or composite measures, such as the ratio of haft width to maximum width related to core shape, core preparation or rejuvenation (if any), the number of platforms, the characteristics of flake scars (including directions of flake removal, prominence of negative bulbs), and intensity of core use (Table 11.1).

Archaeologists frequently classify cores by their basic shape, number of platforms, preparation, if any, and the extent to which removals were parallel and narrow (blades) or broad (flakes).

Single-platform cores is a category that encompasses many kinds of cores that share the attribute of having only one platform from which the knapper removed flakes or blades. Some of these are fairly informal, but others are carefully prepared cores used, for example, to produce blades or bladelets of very regular shape and size.

Pyramidal cores are a type of single-platform core, roughly conical in shape. Some very regular, cylindrical pyramidal cores used to produce long, regular blades, especially by pressure flaking, are called **bullet cores** or **polyhedral cores** (Crabtree 1968). These are often prepared cores.

Multidirectional cores are ones that result when the knapper either alternately or periodically changes the platform from which to remove flakes. The flake scars on such a core cross each other at a variety of angles.

Amorphous cores result when the knapper is opportunistic and, as with multidirectional cores, periodically turns the core to use the scar from a previous removal as a new platform. By the time such a core is near exhaustion, it is typically irregular but somewhat spherical in shape.

Prepared cores include any cores that knappers have intentionally shaped to make subsequent flake or blade removals more predictable. Sometimes that only involved shaping the platform; in other cases, the entire core was carefully shaped prior to striking off any blanks. Some pyramidal and other single-platform cores fall into this category, as do the following types.

Opposed-platform cores are prepared cores with two platforms, one at each end of the core, so that the knapper can remove flakes, or especially blades, alternately from each platform. This tends to result in long, parallel, and fairly regular blades, as the ridges left from previous removals guide the wavefront of fracture. One variety of opposed-platform, prepared core, called a **naviform core** because of its boat-like shape, is a distinctive feature of some early Neolithic sites in the Middle East. It allowed Neolithic knappers to create many very regular, parallel-sided blades that served as blanks for a variety of tools.

Levallois cores are prepared, disk-shaped or tortoise-shaped cores that some Old World Middle Palaeolithic knappers created through bifacial flaking around the perimeter. Once they achieved the desired shape of core, they were able to strike a small number of points, flakes or blades from the less convex of the broad faces of the core before discarding it (Boëda 1995; Van Peer 1992). The shape of the core allowed knappers to control flake shape quite well, making this type of core very suitable for producing large, triangular spear points and broad blades.

11.3.4.3 Attributes of Flakes and Blades

Before even beginning to make measurements on flakes, archaeologists need to decide on how to orient them. Typically, this is by reference to the platform (or proximal end) and either axis of the box length or long axis (Fig. 11.6). Competing versions of size attributes include axial length and maximum length.

Many of the attributes that archaeologists measure on unretouched or minimally retouched flakes and blades are ones that are clues to the reduction technique (e.g., hard- or soft-hammer percussion), reduction strategy (e.g., clues in the dorsal scars to the use of multiple platforms), and intended product (e.g., parallel blade blanks, pointed flakes). They typically include observations at multiple scales (Table 11.1), with flake form and termination type (Fig. 11.7) on a nominal scale but various size and shape indicators on a ratio scale. Some shape indices, such as "pointedness," are ratios of two ratio-scale measures.

11.3.4.4 Attributes of Retouched Tools

Many of the attributes that archaeologists measure on tools are clues to the tools intended function or actual use, such as notching to facilitate attachment to hafts, or steepness of retouch (backing) of a long edge to strengthen the edge for scraping tasks or to dull it to avoid damaging the user's hand or hafting material. Others, like invasiveness of retouch, are measures of how much effort the knapper made to modify the tool's shape and thickness. Still others are clues to the

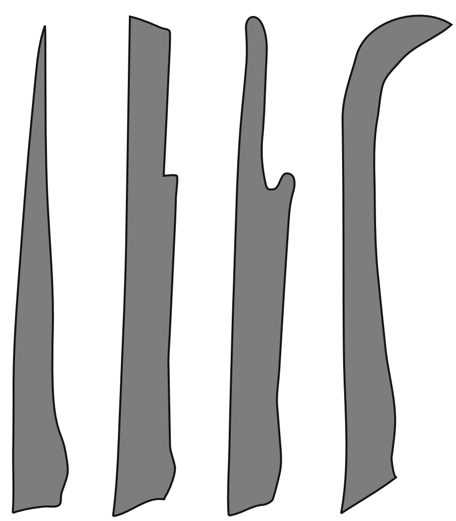


Fig. 11.7 A classification of flake terminations as viewed in radial section: from left to right, feather, step, hinge, and plunging terminations

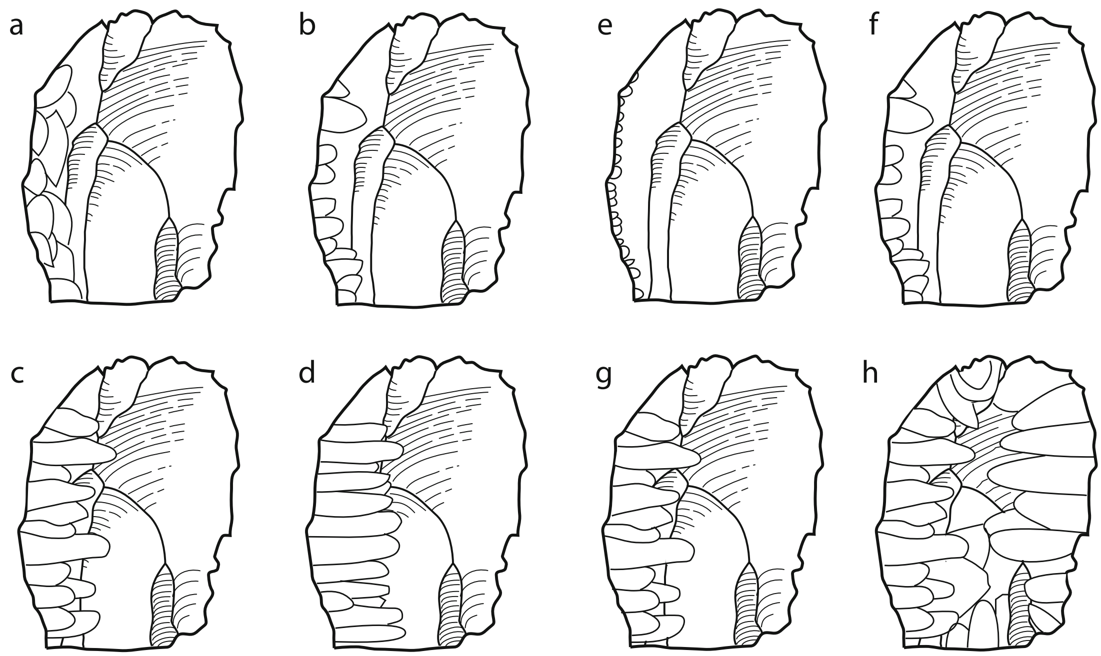


Fig. 11.8 Classification of retouch type and ordinal scale of invasiveness: (**a**) scaled, (**b**) stepped, (**c**) sub-parallel, and (**d**) parallel retouch; (**e**) short, (**f**) long, (**g**) invasive, and (**h**) covering retouch

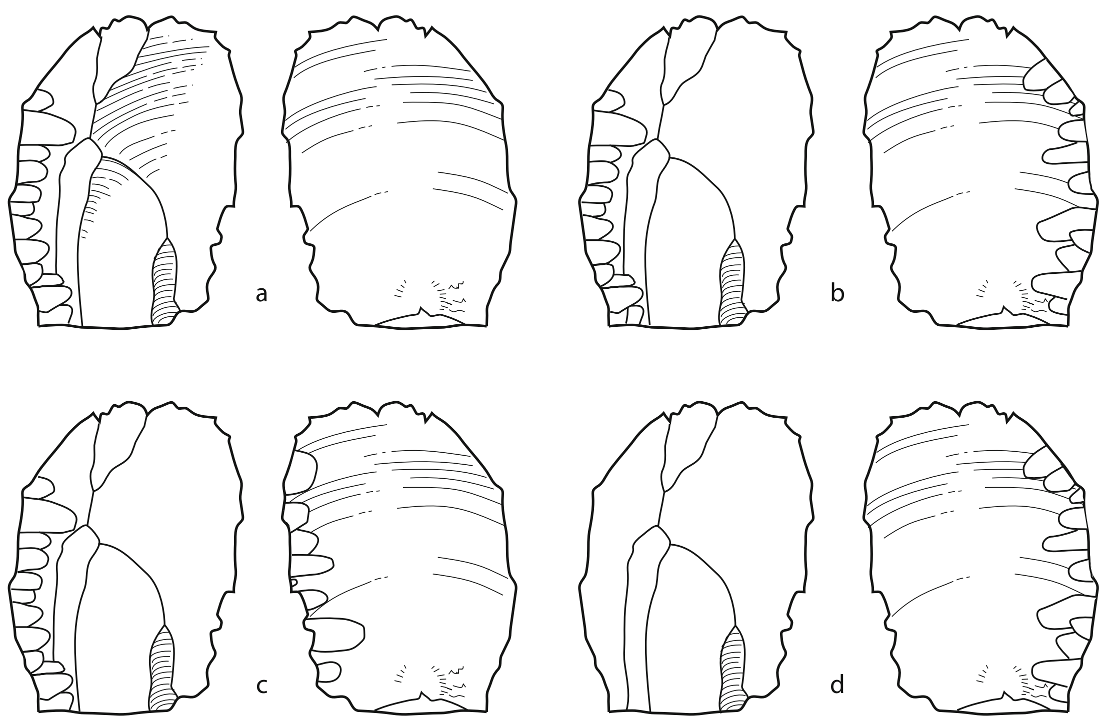


Fig. 11.9 A classification for description of retouch position: (**a**) direct, (**b**) bifacial, (**c**) alternate, and (**d**) inverse

character of the blank that the knapper selected for retouch into a tool.

Archaeologists typically record many aspects of retouch on nominal or ordinal scales, including retouch type, location, and invasiveness (Figs. 11.8 and 11.9), although the last can also be measured on a ratio scale (Hiscock and Tabrett 2010).

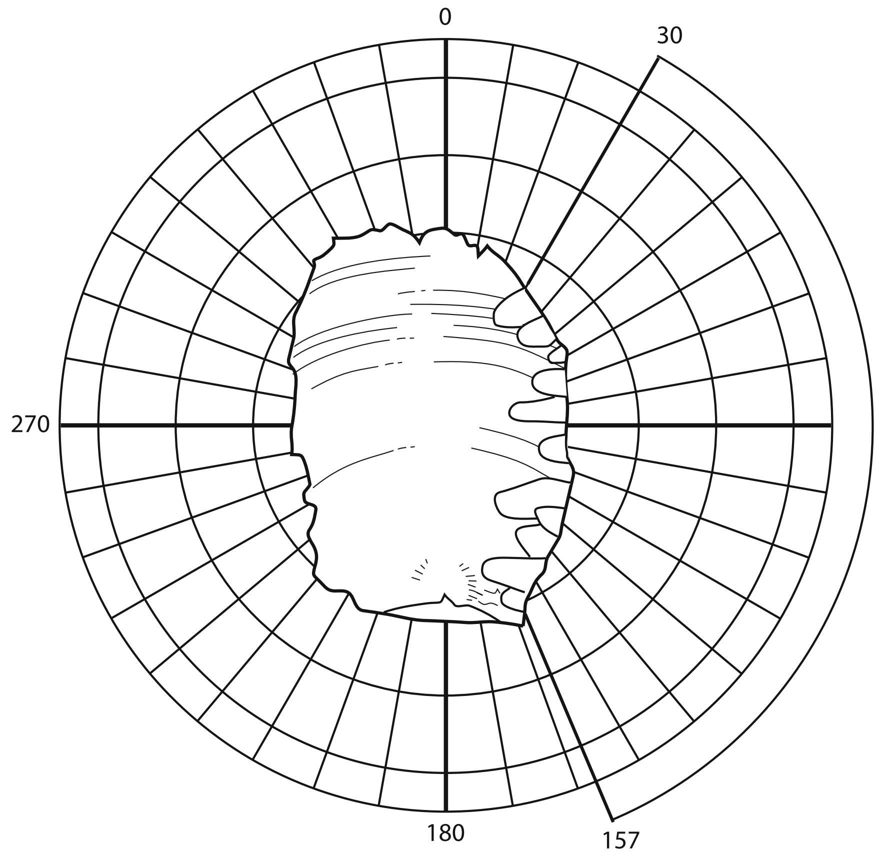


Fig. 11.10 Measuring the "length" of retouch on an edge by the angle of arc in degrees on a polar graph, here 127°. A similar scale can be used to record the locations of polish or use wear. The artifact is positioned here by its axis of flaking (axial length) and midline

Steepness of retouch is the angle between the flat plane of a flake and a retouched surface, as measured with a goniometer on a ratio scale. As it is likely to vary noticeably along an edge, it is important to measure it at intervals along the edge to obtain a mean or trimmed mean and error estimate.

"Length" of a retouched edge could be measured on a ratio scale in millimeters, but this would vary with the overall size of the artifact. When closely similar tools vary substantially in size, it can be better to measure this "length" with the angle of the arc (centered on the artifact) that encloses the retouched area (Fig. 11.10).

A similar approach is also useful for recording the locations of retouch or use wear, sometimes called the **polar coordinates** method. Since retouch can occur in several locations on the same tool, it is important to put careful thought into the "retouch location" field of any database for recording it (see Chap. 4).

11.3.5 Flintknapping Waste

The waste products of flintknapping can be quite important to our understanding of the technologies involved in tool production and maintenance. Because debris includes chips and chunks that, by definition, have no identifiable ventral surface, we cannot analyze them in the same way as flakes and may not be able to identify their manner of fracture (Inizan et al. 1992: 85). It has been common to distinguish "primary" flakes (ones that exhibit cortex over nearly all their dorsal surface) from "secondary" and "tertiary" flakes that have less or no cortex (Jeter 1980), as these would presumably be removed at different stages of core reduction, although definition of these classes has not been consistent. Waste also includes broken flakes and flakes that resulted from knapping errors; these may exhibit some of the attributes that we usually record on flakes and blades – for example, we can measure platform shape on proximal portions of flakes, and termination type on distal portions. However, the sheer volume of waste material in some archaeological contexts has led some archaeologists find fast ways to sort waste products.

One such approach with a long history (Sullivan and Rozen 1985) is a simple taxonomy with dichotomous distinctions. It starts (Fig. 11.11) with distinguishing products with a ventral surface from ones that do not; this separates flakes and blades from chips and chunks. For the former, it distinguishes ones with a preserved point of applied force (intersection of bulb and platform) from ones that do not; this separates complete flakes and proximal flake fragments from medial and distal flake fragments. For the former, we can further distinguish ones that have intact margins from ones that to not. Flakes and blades with complete margins exhibit a hinge or feather termination and both

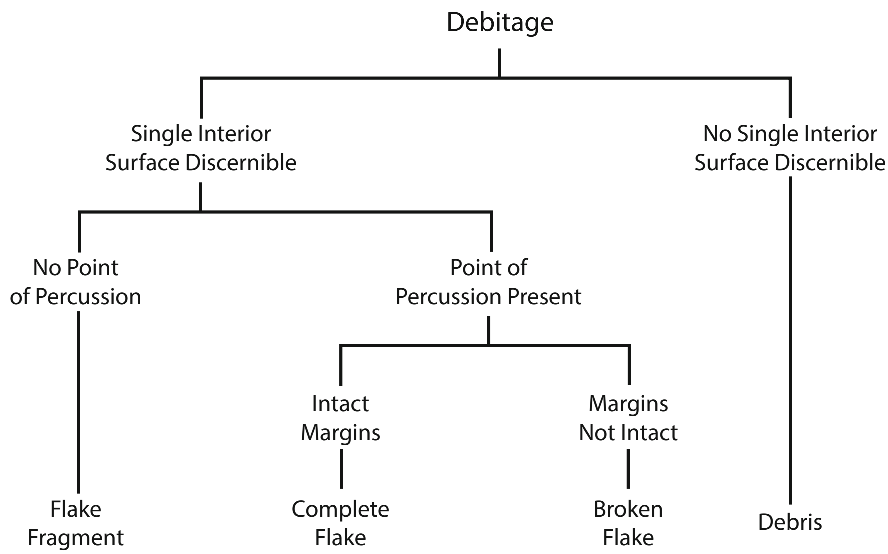


Fig. 11.11 Taxonomy of debitage. (Modified from Sullivan and Rozen 1985)

left and right edges (so that we can measure width). Other archaeologists have adapted this taxonomy to subdivide debris further into chips and chunks (distinguished by size) and flake fragments into medial and medial/distal fragments.

Another method that offers ease and cost-effectiveness is mass analysis (Ahler 1989; Stahle and Dunn 1982). This involves sorting large amounts of debitage into size categories, typically by sieving through nested screens, much as geomorphologists infer depositional processes from particle-size distributions (see also Chap. 17). For each assemblage, "weighing" the debitage caught on each screen produces a size distribution by mass. Interpretation of the distribution depends on analogies to "control groups" – simulated assemblages created by modern flintknappers producing particular types of tools – that are similarly size-sorted.

While such approaches to debitage analysis are cost-effective, they also present several threats to validity when used to infer reduction strategies (Morrow 1997). Both Sullivan and Rozen's and Ahler's approach are susceptible to the risk that an assemblage could be a mixture of debitage from two or more flintknapping episodes, production activities, or flintknappers pursuing different reduction strategies or tool types. Various processes may also have differentially removed some of the debitage originally deposited in the assemblage, including, not surprisingly, selection of blanks for tool production. Nor do these methods typically account for differences in raw material (Andrefsky 2007). In the case of mass analysis, confounding factors could also affect the character of the "control groups" (see pp. 91–94), while it is also not well demonstrated that size distribution is the best way to distinguish technologies.

Consequently, these methods have mostly given way to analyses that highlight evidence for technological decisions and actions during flintknapping. Such evidence includes types of flakes and other products that are specific to particular reduction processes, such as correcting knapping errors or rejuvenating cores (e.g., core tablets), thinning bifaces, or truncating blades. The presence of any number of such products tells us something about lithic reduction even when much of the production is missing from the assemblage. The evidence also includes attributes of waste flakes and discarded cores and core fragments that are strongly associated with hard- or soft-hammer percussion or pressure flaking, or with specific strategies of core reduction, such as direction of previous flake removals. This shift is also partly due to the growing role of the *chaîne opératoire* in lithic studies, with its emphasis on the whole technological process in its social context, and not just on final products.

11.3.6 Refitting

Some lithic assemblages preserve enough debitage from a single core to allow refitting of flakes to reconstruct much or all of the reduction sequence, usually with a few flakes missing because they were used as tool blanks (Fig. 11.12). This not only helps us understand reduction strategies and the criteria for blank selection, it can also help us understand site-formation processes. For example, sometimes debitage from the same core is found in different spatial (Fig. 11.13) or stratigraphic contexts (e.g., Deschamps and Zilhão 2018), potentially showing patterns in terms of which flakes were selected for use, while others were left at the knapping location. In ideal circumstances, refits can also be clues to mobility strategies, when unmodified debitage from one site refits to tool blanks at another site, or when flakes missing among the refitted debitage from the first site appear to have been used for tools found at the second site.

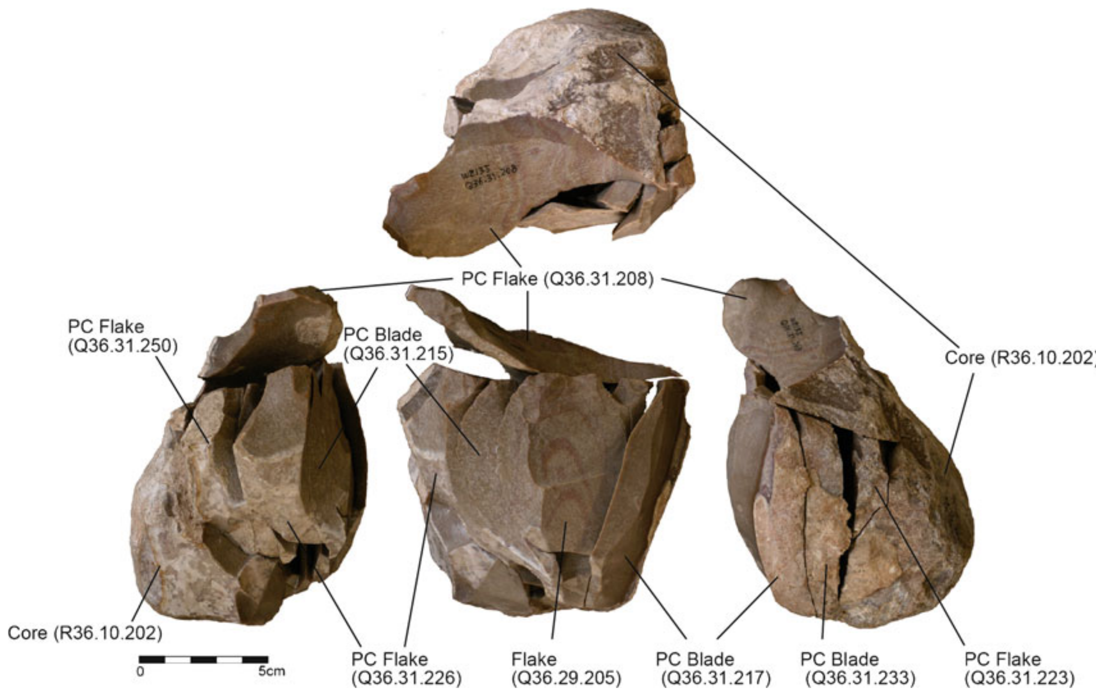


Fig. 11.12 Flakes refitted to a core from the Late Neolithic site of al-Basatîn, Jordan (From Kadowaki and Banning 2018: 68). "PC" is partially cortical

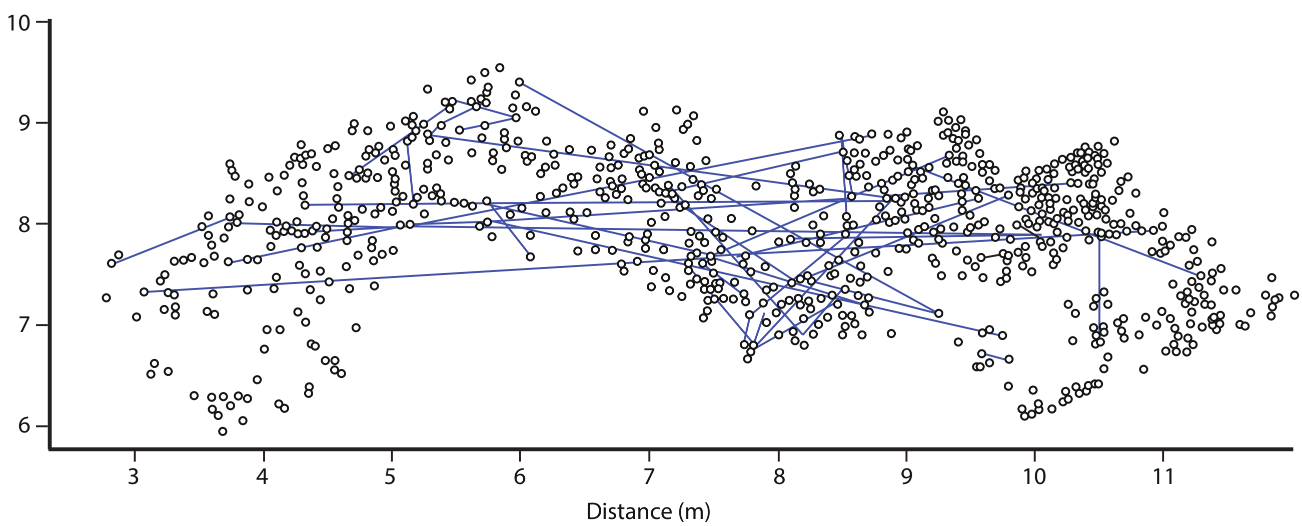


Fig. 11.13 Map of connections between refitted lithics in Layer 15 at Oliveira Cave, Torres Novas, Portugal. (Modified from Deschamps and Zilhão 2018)

11.4 Ground-Stone Tools

Although flaking is one of the techniques used to shape many kinds of ground-stone tools, their main characteristic is a somewhat smooth surface that results from grinding or polishing, either during use or as part of the manufacturing process.

11.4.1 Manufacture of Ground-Stone Tools

Ground-stone tools include not only grinding stones, such as querns, mortars, pestles and millstones, but also axes, adzes, hammers, knives, mace-heads, sling stones, stone vessels, smoking pipes, palettes, mirrors, sculptures, beads, spindle whorls, shaft-straighteners, standard weights, and other tools or ornaments made or finished by pecking, drilling, grinding and polishing, and even hammerstones that are only very slightly modified natural pebbles or cobbles (Adams 2002; Rowan and Ebeling 2008).

In selection of raw material for this wide variety of tools, fracture mechanics is typically less important than other characteristics, such as hardness, and whether it was desirable for the finished product to be abrasive (for milling stones) or smooth (e.g., for adzes or decorative items). Most raw materials used for ground-stone tools were relatively unsuitable for conchoidal fracture, although flaking could still be used to rough out a blank to be finished with other methods.

After selection of raw material, manufacturers typically used a mix of reductive and finishing techniques to quarry stone and shape blanks (Lewis et al. 2011; Stocks 2003). Reconstructing the reduction sequence is more difficult than for flaked-stone tools because later manufacturing steps typically remove most or all traces of the early steps in the process, and waste from manufacture is not always very informative. In some cases, however, quarry waste can provide clues to the early stages of manufacture (Huckell 1986; Nelson 1987; Schneider et al. 1996).

Fire-cracking is a method that is useful for removing blocks of stone from a quarry. It involves cutting or chipping out grooves, or drilling rows of holes, around the intended block, building fires along these indentations, and then pouring cold water on the heated grooves or holes so that the sudden temperature drop will cause cracks to propagate and the rock to split (Barber 2010: 71–72).

Wedging and leverage are similarly useful for snapping blocks from their place in a quarry. Beginning with the same grooves or holes as in fire-cracking, stoneworkers hammer wooden or metal wedges into them, preferably with several being hammered simultaneously and with rhythm, or large levers of metal or dense wood are inserted and then sharply pulled simultaneously. Either method can snap a block along the desired line, especially when quarrying layered rocks like limestone that already have strong natural bedding planes (e.g., Wagalawatta et al. 2016: 170–172). These methods can also be used to remove the central column left by tubular drilling (see below; Stocks 2003: 134–135).

As noted, tool makers could use flaking to rough out a blank, making much the same use of conchoidal fracture that a flintknapper would, but with less precision. Stone masons also used flaking to trim stones for constructing walls.

Pecking, striking the stone with a hammerstone at something close to a 90° angle, can also help rough out tool shapes and remove irregularities in the surface (Wright 1992: 55–57).

Chiselling, especially when metal chisels are available, works more quickly than pecking as the sharp chisel edge can be struck against the stone at a more acute angle, using indirect percussion with a mallet.

Scraping with flint scrapers can be effective for further shaping tools or vessels made from softer stones, such as gypsum or calcite. Traces of the scraping can later be removed by grinding or polishing.

Cutting or **sawing** can be used to split larger slabs of material, to achieve desired dimensions, and to correct faults. Prior to the invention of diamond saws, sawing depended mainly on the availability of metal saws (e.g., Stocks 2003: 32), which were used most commonly on relatively soft stones. Unaided, copper and bronze saws are too soft to be effective for sawing very hard stones, but combining the saw with an abrasive, such as wet quartz sand, makes it possible to saw even granite (Stocks 2003: 108–109, 116–118). Use of such abrasives makes it possible to saw many kinds of stone even with string. Sometimes, traces of the abrasive persist in sawn channels and, if they are stained with copper residue, this indicates the use of copper saws.

Grinding involves removing material from a blank by rubbing it aggressively with a hard, coarse stone, sometimes aided by sand and water, in either a circular or back-and-forth motion. It is particularly useful for creating flat, fairly smooth surfaces on grinding stones, sculptures and building stones (Dickson 1980: 162–163). While the grinder is usually a portable tool, in other cases the toolmaker may rub the unfinished artifact against an abrasive slab of stone (e.g., Andrieu et al. 2014). Once hard metal files were available, filing could also be used to remove raised areas on blanks.

Incising involves using a hard, sharp edge and a longitudinal motion to remove material from the blank to create a long, rather straight groove, often with a V-shaped section. This groove can be a finished feature or can delineate areas on the blank that will be further removed by pecking, drilling, or chiselling (Wright 1992: 55–57).

Drilling stone can involve using flint drill bits and a bow drill, and the resulting perforations are recognizable as "biconical," as it is necessary to drill from both sides with the narrowest part of the perforation where the two drill holes meet. However, it is also possible to drill nearly cylindrical holes in stone with a tubular drill made from reed or cane and wet sand as an abrasive, as was common in ancient Egypt (Gwinnet and Gorelick 1987; Stocks 2003: 104–105, 111–112). A combination of drilling and boring can be used to hollow out stone vessels.

Early drills only had a capstone to act as a bearing on the upper end of the shaft but the discovery that the shaft could be fixed with bearings at both ends led to the development of the cutting wheel. A hard, sharp disk rotated on the shaft could then be used to cut out linear or curved areas on beads or cylinder seals (Sax and Meeks 1995; Sax et al. 2000).

Like drilling and wheel-cutting, lathes use rotary motion to remove material from stones. Typically, the blank is mounted horizontally on a lathe and rotated with foot treadles to provide kinetic energy, while the manufacturer holds a chisel or disk-shaped tool against the blank's rotating surface to grind it down (Sax et al. 2004).

Polishing by rubbing the surface of a nearly finished tool, bead, or vessel with something like a leather lap and a fine, wet abrasive, such as silt, loess or carborundum, can remove small irregularities and striations and result in a glossier surface (Sax et al. 2004).

11.4.2 Anatomy of Common Ground-Stone Tools

The immense variety of ground-stone tools makes a unified information language rather overwhelming, so this section concentrates on some of the most common types of grinding stones: lower milling stones (querns, *metates*), upper milling stones (handstones, *manos*), mortars and pestles (e.g., Wright 1992). Other tool types require quite different terminology, some of it shared with aspects of pottery or flaked-stone tools (Adams 2002).

One subset of ground-stone tools tends to be used in pairs for pressing, pounding, crushing, pulverizing, or grinding seeds, minerals, or other materials. One member of the pair is a relatively fixed, lower element with its use-surface upwards (e.g., quern or mortar), and the other is a more portable element with at least one use-surface downward, which a user moves up-and-down or back-and-forth repetitively or in a circular motion against the use surface of the lower tool (Biskowski 2008; Carelli and Kresten 1997). Generally, handstones work with lower milling slabs, and pestles with mortars. Some more sophisticated milling equipment includes disk-like millstones of which the upper one rotates, as in recent flour mills, and near-cylindrical upper stones that roll in a tight circle against a lower stone, as in some olive presses.

Most lower milling stones (querns) are relatively large, fairly flat, and typically elongated, with a table-like or somewhat concave upper grinding surface. During use, the quern's long axis would be oriented away from the user, who would lean forward to push a handstone, oriented at right angles to the quern, longitudinally in a back-and-forth motion. Frequently, the proximal end of the quern is thinner and the use surface is more concave near the distal end (Fig. 11.14). The upper, use surface is the ventral side, and the lower surface is the dorsal side.

The handstone used with the quern is typically ovoid, domed, or bun-shaped, with a flat or slightly convex use surface on the bottom, ventral side, and a very convex dorsal side to provide a grip for the user's hands (Fig. 11.15). However, some handstones have two use surfaces, one on either side, making it impossible to distinguish ventral and dorsal sides in a meaningful way, although we might

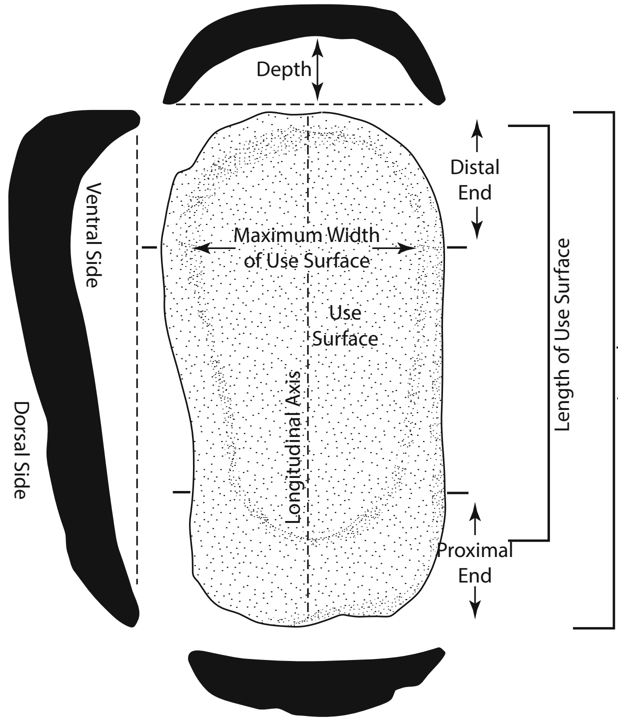


Fig. 11.14 Segmentation terminology for a lower milling stone or quern. (After Wright 1992: 497–498)

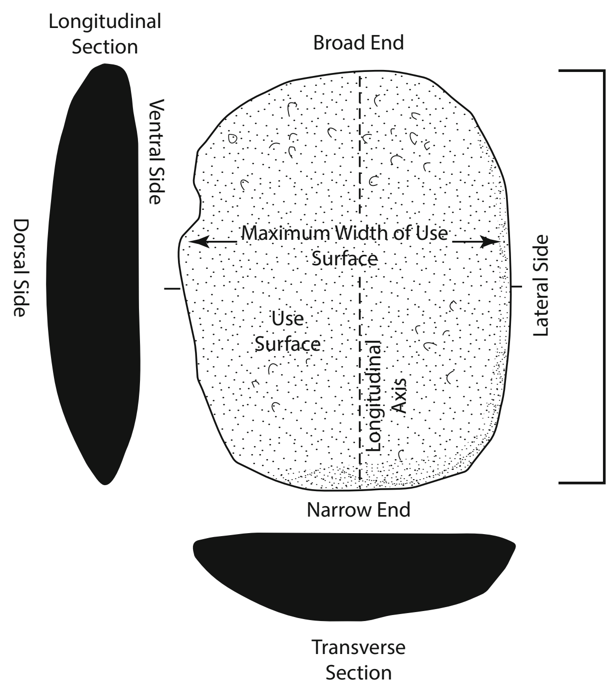


Fig. 11.15 Segmentation terminology for an upper milling stone or handstone. (After Wright 1992: 497–498)

arbitrarily decide to designate the most convex side as dorsal. Typically, handstones are somewhat elongated, to facilitate gripping with two hands, but it is impossible to distinguish proximal and distal ends except arbitrarily, and it would have been common to alternate the orientation of the tool to prevent uneven wear. In some cases, handstones are circular, making such distinctions completely impossible.

Mortars are lower stones that, because they are typically circular in plan view, cannot have proximal or distal ends. The ventral or upward use surface is in a highly concave opening, often giving mortars a very bowl-like shape (Fig. 11.16) and may include the entire concavity or only its lower portion, as indicated by battering or greater wear. Their base, analogous to the dorsal surface of a quern, is typically flat or slightly concave to encourage stability during use but can be somewhat convex. Some mortars are quite deep, with nearly vertical side walls, while others are squat. The exterior of mortars can be regular and carefully made, but sometimes shows the surface of the unmodified stone from which the mortar was made. It is even common for mortars to be fixed facilities in bedrock, rather than somewhat portable artifacts. Smaller artifacts that technically have the features of mortars may not have been used for pounding or grinding, but as sockets for door-pivots, or cap-stones for bow-drills, and ones with rows of concave depressions could even be parts of games.

Pestles are elongated cylindrical or conical tools with one or two convex use surfaces at one or both ends for pounding or grinding in a mortar (Fig. 11.16). Some pestles with a single use surface may show carved decoration at the opposite end. Only in pestles with a single use surface can we distinguish distal (use) from proximal end, except by arbitrarily counting the narrower end as proximal. Pestles, incidentally, can be made from materials such as dense wood, as well as stone.

Ground stone and ground edges were also prominent in an array of hafted pounding and chopping tools, such as axes, adzes, mace heads, and hammers (e.g., Dickson 1976; Geneste et al. 2012; Latorre et al. 2017; Lewis et al. 2011; Takashi 2012). Axes and adzes no doubt were important for felling trees and trimming timber, and potentially also as weapons and status symbols, while stone hammers were important for pulverizing ores for metal extraction and other purposes.

Pendant-like, elongated items of ground stone could be pendants for personal adornment, but they could also be whetstones for sharpening metal tool edges (Fig. 11.17), touchstones to test the purity of metal alloys by the color of a streak (Jezek 2013), and even bull-roarers ("aerofoils" swung on a string or thong as musical instruments).

11.4.3 Some Attributes of Ground-Stone Tools

As usual, the attributes that archaeologists measure on ground-stone tools depend on their research questions, but typically include ones related to function, scale of processing, source of raw material, motion of use, and material cut, crushed or pulverized (Table 11.2).

Because pounding and grinding are abrasive processes, we can expect pronounced use wear on many kinds of

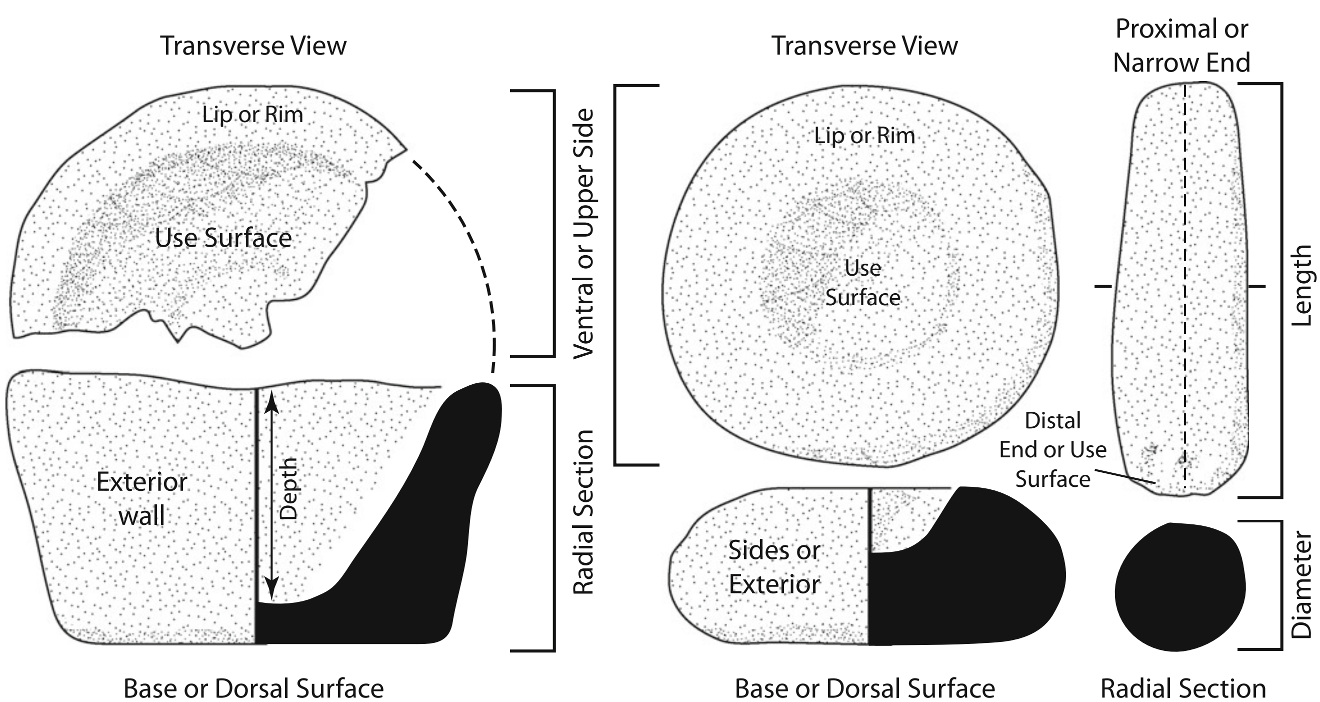


Fig. 11.16 Segmentation terminology for mortars and pestles. (After Wright 1992: 497–498)

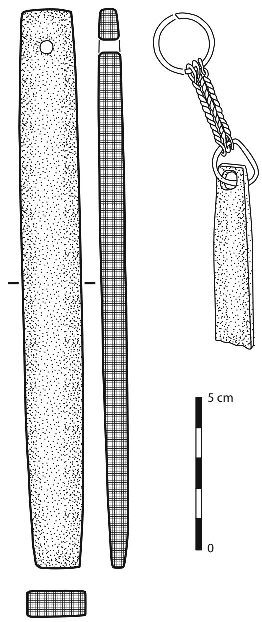


Fig. 11.17 Tenth- or Eleventh-century whetstones from Starigard-Oldenburg, Germany. (Y. Salama, after Gabriel and Kempke 2011)

ground-stone tools (Adams 1988, 2002: 25–42). However, some kinds of wear or polish from use can be difficult to distinguish from design features and abrasion during manufacture.

11.5 Use-Wear and Residue Analysis

Considerable archaeological research focuses on the traces that occur on stone tools as a result of their use (Marreiros et al. 2015; Stemp et al. 2016). This is particularly well-developed for chipped-stone tools but is also applicable to ground-stone tools (e.g., Adams 1989, 2002; Bofill 2012; Dubreuil and Savage 2014; Latorre et al. 2017).

The pioneer of lithic use-wear analysis was the Russian prehistorian, Semenov (1964), who realized that microscopic scratches and polish on the edges of stone tools are evidence of the materials with which the tools came into contact, and the tool-users' motions.

Table 11.2 Examples of attributes for selected ground-stone tools with typical scale of measurement and some possible purposes or indirect measures

Attribute	Scale	Purpose
Querns or lower milling stones		
Length of use surface	Ratio	Length of user's stroke
Area of use surface	Ratio	Scale of production
Concavity of use surface	Ratio	Volume, scale of production, use life
Texture of use surface	Nominal, ordinal or ratio	Function, wear management, or material ground
Orientation of use wear (rotary, longitudinal, etc.)	Nominal	Function
Presence or type of residues	Nominal	Material processed
Handstones or upper milling stones		
Shape of use surface	Nominal or ratio	Motion(s) of use
Mass	Ratio	Intensity of use
Curvature of use surface	Nominal, ordinal or ratio	Motion(s) of use
Use-surface texture	Ordinal	Wear-management
Macroscopic use traces or residues	Nominal	Material processed (e.g., ochre), motion(s) of use
Mortars		
Ratio of interior rim diameter to basin depth	Ratio	Motion(s) of use; duration of use
Interior volume	Ratio	Scale of production
Orientation of use wear (rotary, vertical, etc.)	Nominal	Motion(s) of use
Residues of lipids, starch, phytoliths	Nominal	Material processed (e.g., nuts)
Condition of base	Nominal	Complete or pierced from extensive use
Pestles		
Number of use surfaces	Nominal or ratio	Motion(s) of use
Mass of complete pestles	Ratio	Crushing force, Scale of production
Striations or chipping on use surface	Nominal	Motion(s) of use
Axes and adzes		
Location of polish	Nominal	Function
Extent of polish	Ordinal	Duration of use
Mass of complete tools	Ratio	Function, resharpening
Length of bit edge	Ratio	Function
Edge angle	Ratio	Function, resharpening
Cross-sectional edge symmetry	Ordinal	Function (axe or adze)
Edge chipping	Ordinal	Function, resharpening, duration of use

Most use-wear analyses of chipped stone focus on two main dimensions: motion of use and contact material. For the former, analysts try to identify fractures, scratches, striations, and other attributes that allow distinction of longitudinal, transverse, and rotary or drilling motions, as well as damage from impact (e.g., on projectile points). For the latter, they focus mainly on aspects of polish that may be associated with use of the tool on different materials, such as meat, animal hide, bone, antler, wood, and grasses. Some of the attributes of polish are brightness, pitting, striations, and troughs. Effective use-wear analysis also has to deal with the possibility of post-discard and post-depositional damage that is not related to use.

The methods employed in use-wear analyses include both low- and high-power microscopy with optical light, scanning electron microscopy (SEM), and, more recently, surface metrology, each of which has different strengths. Low-power optical microscopy, in the range of 20-50x, can be effective for identifying fractures and the most highly visible polishes (Fischer et al. 1984; Grace 1989, 1996; Vaughan 1985), but high-power microscopy, in the range of 100–1000x, is more effective for examining polishes and fine striations, but is more time-consuming and has low depth of field. SEM has greater depth of field and high magnification but is best used in combination with optical light microscopy (Borel et al. 2014). Surface metrological methods include laser-scanning confocal microscopy (Evans and Donahue 2008), interferometry (Anderson et al. 2006), laser profilometry (Stemp et al. 2009), atomic force microscopy (Faulks et al. 2011), and focus-variation microscopy (Macdonald 2014). These last methods allow greater differentiation of polishes on lithics by processing the measurements with algorithms that provide quantitative data on the texture, or surface roughness, of polished areas. In all cases, it is necessary to clean the lithics gently with a mild detergent or chemical wash to remove adhering dirt and oils while avoiding alteration of the use traces.

Although recent ethnographic observations of indigenous people who used stone tools provides useful clues, a key feature of use-wear analysis is that it depends on experimental controls (Stemp et al. 2016). We identify the use-motions and contact materials by analogy between the traces we see on archaeological lithics and those we see on experimental lithics whose uses and use motions are known. To do this effectively and with validity, the experiments should account for, not only the factors of motion and contact material, but all the likely confounding factors that might affect the character of use wear. These could include raw material and edge morphology of the tool, skill, fatigue and strength of the user, angle between the edge and the contact material, humidity of the contact material, and others. Furthermore, while it is tempting to control for these one at a time, it is likely that some of them interact, so that a factorial design is preferable (see Chap. 6). In addition, it is advisable to use blind or double-blind testing (Evans 2014) and random mixing of lithics from each group (experimental and archaeological or ethnographic, and different motions and contact materials) to prevent analysts' preconceptions from jeopardizing the validity of the results. Study participants should not know in advance to which group each lithic they examine belongs.

Recently, residues on stone tools have been increasingly important in inferences about their use. Studies of blood residues, lipids, and starches have been particularly notable (Anderson 1980; Anderson et al. 2006; Barton 2007. One of the earliest examples of lithic residue analysis involved detection and identification of blood antigens on tool edges (Lowenstein 1985). Early analyses were expensive and often problematic in archaeological contexts, but enzyme immunoassays (EIA) and crossover electrophoresis soon made identification of blood residues cheaper and more applicable to archaeological specimens (Hyland et al. 1990; Kooyman et al. 1992).

11.6 Economizing Behavior and Design in Stone Tools

Lithics analysts, especially those favoring an economic-materialist perspective, have long shown interest in tools' design features (Bamforth 1986; Horsfall 1987). This involves assuming that hunters, farmers, fashioners of groundstone tools, and flintknappers made decisions based on their rational evaluation of known alternatives that had different costs and benefits as measured in time, energy, effort, risk, or some other "currency," and manipulated competing factors in their attempts to achieve a pre-conceived design. In some hunting strategies, for example, it may be advantageous to have weapons that are highly reliable, while other strategies might call for weapons to be easily and quickly maintainable (Bleed 1986; Hughes 1998; Hutchings 1991; Knecht 1997).

Lithic analysts have often observed variations in assemblages that they interpret as evidence for increases in the economic efficiency of tool makers. For example, Mesolithic flintknappers in Eurasia and Late Stone Age knappers in Africa were able to get many meters of usable edge from the same nodule of flint that a Lower Palaeolithic or Early Stone Age knapper would have used to obtain only 30-40 cm of cutting edge. Patterns such as these may suggest that knappers who produced mainly very small flakes or blades to use as tool blanks were practicing economizing behavior by get more useful tools from the same volume of core. Evidence for repeated re-sharpening, reworking or re-use of tool components can also suggest economizing behavior, as can the manufacture of tools that tend to have several distinct usable edges with different edge angles, so that the same tool can be used for different tasks (Vierra 1995). These are arguably the "Swiss Army knives" of prehistory.

One theoretical perspective that a few lithic analysts have adopted is Design Theory (Lindner and Rodger 2009; Pye 1964). According to this engineering-based theory, every technological design constitutes a compromise between cost, effectiveness, risk and reliability, and, like other economizing theories, there is an assumption of intentionality and pre-conceived design. Large investments in design, even over-design, appear in contexts where the consequences of tool failure are unacceptable, so that there are back-ups or redundant systems to ensure reliability. Inexpensive, *ad hoc* or more casual tools with little investment in design are associated with tasks that have low risk (Bamforth and Bleed 1997; Banning and Siggers 1997).

However, "wasteful" reduction strategies, such as bipolar reduction to create many rather randomly shaped flakes only some of which are selected as tool components, may suggest "expedient" technologies where raw material is abundant and not particularly valued, or where such expedient tools are simply "good enough" for the task at hand. In some of these technologies, there may be very little pre-planning of lithic forms, or none at all, the knappers simply producing a large number of flakes from which they select ones that have potential. In such technologies, retouch actually represents, not preconceived design, but attempts to correct unsatisfactory features, and may even lead to rejection of the piece (Hiscock 2004: 75). This implies that retouch is not necessarily a good indication of a design concept.

11.7 Style in Stone Tools

Although intended function, available technology, and raw materials place limitations on how a flintknapper makes a tool, there are still many ways to accomplish the same task that vary either substantially or in detail. Some archaeologists use "style" to describe any variation in a tool that they cannot explain by variation in material, technology, or function (Close 1978), others as the sum of technological choices and technical acts (Lechtman 1977). They may assume that gross aspects of style are passed from generation to generation, in gradually distorted form, by toolmakers teaching others their craft (cf. random copying, Bentley et al. 2004; Shennan 2002). Yet archaeologists disagree on how to conceive of style. Sackett (1977, 1982, 1990) offers an influential model for what he calls **isochrestic style**. He suggests that there are several ways to make a tool that would satisfy its function, so that the form archaeologists find represents the toolmaker's choices among these alternatives in a particular context. We presume that this context included the traditions, knowledge, habits, and values of the person or group that made the tool, so the selection among functionally equivalent alternatives was restricted. Thus, we would not expect two social groups widely separated in space and time to exhibit the same choices, while the degree of similarity between interacting groups would depend on their degree of social interaction. In this respect, Sackett's isochrestic style brings to mind the *chaîne opératoire* and communities of practice, except that Sackett focuses on the form of final products while others would emphasize choices and technical acts (Lechtman 1977; Lemonnier 1986: 148). As in evolutionary theory, selection can act on these variations, so that some became more common while others disappeared.

By contrast, Sackett would call more deliberate use of style to signal group membership and other information symbolically, **iconic style**. Sackett (1990) notes that many researchers seem to restrict iconic style to "adjunct" aspects of form, such as the decorative or non-instrumental aspects, but iconic information can also be latent in the isochrestic choices. Archaeologists who make use of design theory or evolutionary approaches often favor some version of the isochrestic model (e.g., Sheppard 1987; see also Shennan 2009).

However, some ethnographic research indicates that style is not so passive. A high degree of social interaction, rather than leading to stylistic similarity, sometimes leads to stylistic divergence, as people, intentionally or not, emphasize social boundaries (Davis 1990; Hodder 1982; Wobst 1977). Wiessner (1983, 1985, 1990) suggests that style constitutes a premeditated behavior to communicate social messages. She considers style to be variation that results from the human cognitive process of comparing styles and social identities. She does agree with Sackett that style occurs in both decorative and functional attributes, and calls styles with very distinct symbolic referents **emblematic style**, while those with only vague associations with social identity **assertive style**.

Other researchers have focused on the most idiosyncratic levels of style, those apparently due to individuals' differing motor habits and abilities (Hill and Gunn 1977). This involves methods similar to those of handwriting analysis or detection of art forgeries in that they identify the output of individuals or workshops rather than ethnic or other large social groups, although this approach, when paired with *chaînes opératoires*, also grades into communities-of-practice research.

That is because the *chaîne opératoire* emphasizes gestures and sequences of gestures, as embedded in learned cultural behavior, so that it overlaps in some respects with research on style (White 1993).

In addition, some stylistic aspects of lithics, iconic in Sackett's terms, can be deliberately symbolic. Some very carefully made stone tools were never intended to be used for utilitarian functions, such as cutting or chopping, and some were even so finely made and so thin that they would surely break if anyone attempted to use them. Among these we might count Mayan "eccentrics" that are impressive examples of a flintknapper's craft but have no obvious utilitarian purpose (Hruby 2007), and some European and Asian axes that are probably symbolic battle axes, but too fragile for actual use in battle.

11.8 Non-Use Alteration of Lithics

Aside from damage caused during the originally intended use and even re-use of tools, archaeologists are interested in evidence for the disposal and post-depositional histories of tools and tool-making waste.

Burning, for example, can result from the discard of flakes in a hearth or from heat-treatment of chert to improve its flaking characteristics (Flenniken and White 1983). This can create distinctive scars called potlid fractures, crater-like pits that result when marked temperature differences force round, dome-like flakes to spall off, although freezing and thawing can also create similar scars. Burning can also cause color changes, especially when the material has iron minerals in it, and crazing, which looks like a network of small cracks on the surface of the stone. It is also possible to detect heat treatment with methods such as thermoluminescence (TL) electron-spin resonance (ESR), infrared spectroscopy (IR), and measurement of remanent magnetism and magnetic susceptibility (Borradaile et al. 1993; Pavlish 1978; Robins et al. 1978; Schmidt et al. 2013).

Trampling, especially where herd animals may have been present on a site, not only can damage the edges of flakes, but sometimes mimics retouch (Shea and Klenck 1993). Repeated stepping on a flake whose edge rests on a small pebble can even result in notching.

Weathering can have a number of effects on stone tools. One, hydration of the outer "skin" of the flint by humidity in the air and soil, can result in a patina on the surface that has a different color than that of the underlying material (Nadel 1993). Patina can take a long time to form, making it a very crude indicator of age, but the rate of weathering depends on environmental variables, especially temperature. Obsidian hydration dating is a dating method that depends on this rate of diffusion of moisture into the obsidian. Desert varnish is a lustrous surface on the stone that results from the action of lichens or micro-organisms (Watson and Nash 1997: 90–91).

In some contexts, it is likely that intentional destruction of stone tools after their normal use life had symbolic significance. For example, Wright (2014) argues that the relatively high cost of querns at Neolithic Çatalhöyük in Turkey and the probability that they were deliberately broken after a period of use may mean they were shared among households and excluded from inheritance. In Central America, meanwhile, the Maya may have intentionally snapped stone tools in half to release their "life force," consistent with an animistic belief that seemingly inanimate things, not just humans or even animals, have spiritual essence (Jackson 2017; Stemp et al. 2019).

11.9 Validity and Reliability in Analysis of Stone Tools

Lithic analysts have paid at least as much attention to the validity and reliability of their measurements as other archaeologists. They have devoted some of this attention to differences among competing measures (e.g., maximum length vs. axial length vs. box length) in the reliability with which analysts can measure them. Others focus on the validity of measures as indirect measures of some other quantity, such as the validity of platform depth as an indicator of original flake mass (Dogandzic et al. 2015). Still others focus on the problem of inter-analyst differences or intra-

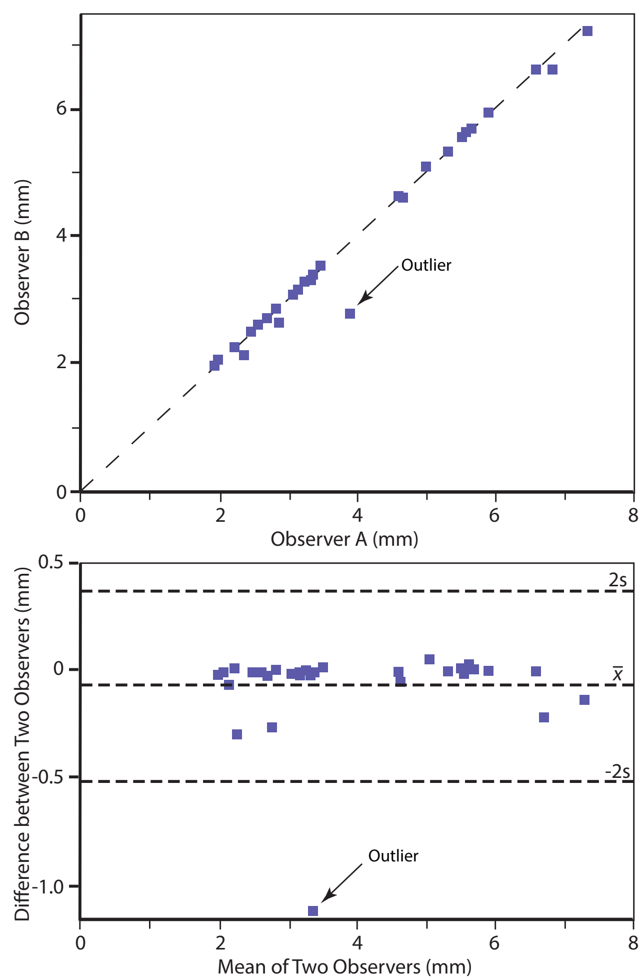


Fig. 11.18 Scatterplot and difference-against-mean plots showing inter-analyst variability for two analysts measuring the thickness of projectile points. (After Lyman and VanPool 2009: 498)

analyst differences over time (e.g., Fig. 11.18; Lyman and VanPool 2009).

As some lithic measurements are somewhat awkward to make in practice – for example, trying to place a goniometer just right against the irregular surfaces of a flake to measure platform angle or edge angle – it is essential to evaluate the likely errors in these measurements and to put protocols in place to ensure that analysts make them as consistently as possible (e.g., Dibble and Bernard 1980). Some common practices, such as multiplying the platform width by the platform depth to estimate platform area, result in considerable bias, as platforms are never rectangular (Muller and Clarkson 2018). One should also be wary of reported values with many significant digits or claims that emphasize small differences in the central tendencies of such measurements, especially in publications that make no reference to measurement error.

Although this technology is not currently as accessible as we might like, the rapidly growing availability of 3D scanning and digital image analysis provides new avenues for reducing the errors in measurement of stone artifacts (Bretzke and Conard 2012; Lin et al. 2010; Muller and Clarkson 2014). Even measurements based on 3D scans, however, can have errors that result from imprecision or inconsistency in orientation or definition, or in the identification of landmarks for defining measurement axes.

Case Study

Role of Earliest Tool-Making in the Evolution of Language

Morgan et al. (2014) used experiments with modern adult subjects to see how the transmission of tool-making knowledge and skill varies among five kinds of transmission ranging from simple reverse-engineering (the knapper figuring out how to knap by intuition and examining previously knapped artifacts) and imitation or emulation to teaching with gestures and verbal language. Their study involved 184 participants who varied in their "distance" along the chain of transmission and 6,000 pieces of flint that they evaluated with six measures of the quality of production. These quality measures included the number of viable flakes produced, the proportion of viable flakes, the number of viable flakes produced per minute, the probability that striking with the hammerstone would produce a viable flake, the expected proportion of core reduced, and the total quality of all flakes.

On all of these measures, imitation and especially reverse engineering yielded rather poor quality, while gesturing yielded results with generally about twice the quality of imitation and teaching with verbal language much greater still (Fig. 11.19). With gestural and verbal teaching, both the probability that a strike with a hammerstone produced a viable flake and the proportion of viable flakes were higher than for other forms of transmission, but declined with distance along the chain of transmission, with all forms of transmission showing similar performance after four or five transmissions (Fig. 11.20).

While they acknowledge that their studies were too short in duration to provide accurate measures of transmission, Morgan et al. (2014: 5–6) conclude that Oldowan tool making would have created selective pressures that favored more complex kinds of teaching, eventually leading to verbal language some time later. They also suggest that imitation or emulation only minimally enhanced the rate of transmission during the Oldowan over reverse-engineering, and this partly explains the apparent lack of change in Oldowan technologies over 700,000 years.

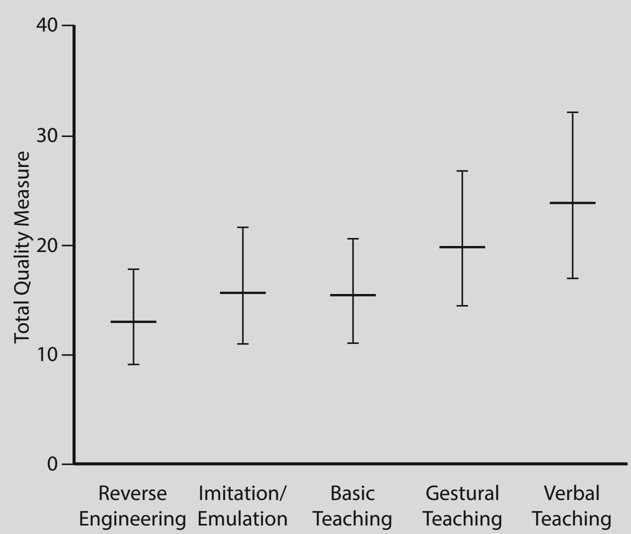


Fig. 11.19 Medians of lithic quality measures for different modes of learning how to strike flakes from cores. Error bars indicate the 95% credible intervals (Bayesian analogue to the Confidence Interval, after Morgan et al. 2014: 4)

(continued)

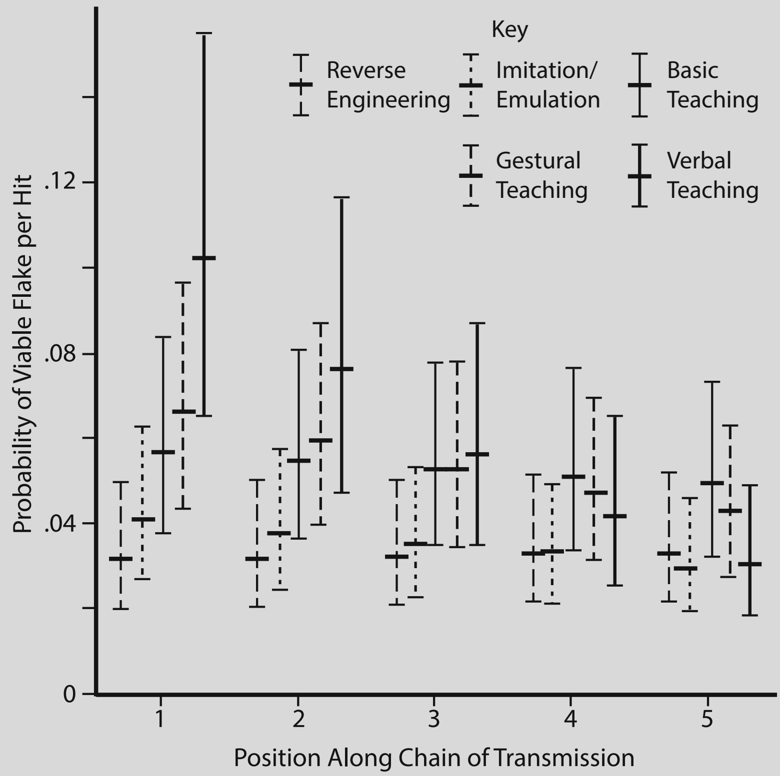


Fig. 11.20 Changes in the probability of striking a viable flake along a transmission chain. For verbal and gestural teaching, this probability declined markedly with distance from the original teacher, so that there was little to no difference among the different kinds of transmission by time 5. (After Morgan et al. 2014: 4)

11.10 Summary

- The *chaînes opératoires* for both flaked- and ground-stone tools include material acquisition, reduction to produce blanks for tools, selection of suitable blanks, further reduction to finish tools, use, re-use, and discard
- Rocks with various properties were differentially selected as raw materials for stone tools, with emphasis on hardness, texture, ease of working, color and aesthetics
- Archaeologists can sometimes discern the sources of raw materials, especially obsidian, by their trace elements or microfossils
- The characteristics of conchoidal fracture dominate the way that knappers can reduce stone cores to flakes, tool blanks, and tools. It can also be important in early stages of the manufacture of ground-stone tools
- Many of the attributes that archaeologists measure on flaked-stone artifacts are ones that help them reconstruct reduction strategies or *chaînes opératoires* of stone-tool technologies
- Refitting flakes and cores is a time-consuming but very useful exercise for helping us understand reduction strategies for flaked-stone artifacts and site-formation processes
- Although some ground-stone tools (axes, adzes) are smoothed versions of flaked stone cutting or chopping tools, others exhibit an array of very different functions, such as pounding, grinding, containment, heat management, weight, or display
- Both classes of tools are subject to use-wear analyses that involve examination of microscopic wear (low-power and high-power), and mineral, chemical or starch residues
- They are also both subject to stylistic analysis.

References Cited

Adams, J. L. (1988). Use-wear analysis on *manos* and hide-processing stones. *Journal of Field Archaeology, 15*, 307–315.

Adams, J. L. (1989). Methods for improving ground stone artifact analysis: Experiments in mano wear patterns. In D. S. Amick & R. P. Mauldin (Eds.), *Experiments in lithic technology 9* (pp. 259–276). Oxford: British Archaeological Reports.

Adams, J. L. (2002). *Ground stone analysis: A technological approach*. Salt Lake City: University of Utah Press.

Ahler, S. A. (1989). Mass analysis of flaking debris: Studying the forest rather than the trees. In D. O. Henry & G. H. Odell (Eds.), *Alternative approaches to lithic analysis* (pp. 85–118). Archeological Papers of the American Anthropological Association 1.

Anderson, P. (1980). A testimony of prehistoric tasks: Diagnostic residues on stone-tool working edges. *World Archaeology, 12*, 181–194.

Anderson, P., Georges, J. M., Vargiolu, R., & Zahouani, H. (2006). Insights from a tribological analysis of the *tribulum*. *Journal of Archaeological Science, 33*, 1559–1568.

Andrefsky, W. (2005). *Lithics: Macroscopic approaches to analysis* (2nd ed.). Cambridge: Cambridge University Press.

Andrefsky, W. (2007). The application and misapplication of mass analysis in lithic debitage studies. *Journal of Archaeological Science, 34*(3), 392–402.

Andrieu, C., Rodas, E., & Luin, L. (2014). The values of Classic Maya jade: A reanalysis of Cancuen's jade workshop. *Ancient Mesoamerica, 25*, 141–164.

Audouze, F., & Karlin, C. (2017). 70 years of "Chaîne opératoire": What French prehistorians have done with it. *Journal of Lithic Studies, 4*(2), 5–73.

Audouze, F., Bodu, P., Karlin, C., Julien, M., Pelegrin, J., & Perlès, C. (2017). Leroi-Gourhan and the *chaîne opératoire*: A response to Delage. *World Archaeology, 49*(5), 718–723.

Ballin, T. B. (2000). Classification and description of lithic artefacts: A discussion of the basic lithic terminology. *Lithics, 21*, 9–15.

Bamforth, D. (1986). Technological efficiency and tool curation. *American Antiquity, 51*, 38–50.

Bamforth, D. B., & Bleed, P. (1997). Technology, flaked stone technology, and risk. In C. M. Barton & G. A. Clark (Eds.), *Rediscovering Darwin: Evolutionary theory and archeological explanation* (pp. 109–139). Archeological Papers of the American Anthropological Association.

Banning, E. B., & Siggers, J. (1997). Technological strategies at a Late Neolithic farmstead in Wadi Ziqlab, Jordan. In H. G. K. Gebel, Z. Kafafi, & G. O. Rollefson (Eds.), *The prehistory of Jordan II. perspectives from 1997* (pp. 319–331). Berlin: ex oriente.

Barber, F. M. (2010). *The mechanical triumphs of the ancient Egyptians* (2nd ed.). London: Routledge.

Barton, H. (2007). Starch residues on museum artifacts: Implications for determining tool use. *Journal of Archaeological Science, 34*, 1752–1762. https://doi.org/10.1016/j.jas.2007.1.007.

Bentley, R. A., Hahn, M. W., & Shennan, S. J. (2004). Random drift and culture change. *Proceedings of the Royal Society B, 271*(1547), 1443–1450.

Biskowski, M. (2008). Maize-grinding tools in prehispanic central Mexico. In Y. Rowan & J. Ebeling (Eds.), *Approaches to old stones – Recent studies of ground stone artifacts* (pp. 144–159). London: Equinox.

Bleed, P. (1986). The optimal design of hunting weapons: Maintainability or reliability. *American Antiquity, 51*, 737–747.

Bleed, P. (2001). Trees or chains, links or branches: Conceptual alternatives for consideration of stone tool production and other sequential activities. *Journal of Archaeological Method and Theory, 8*(1), 101–127.

Blegen, N. (2017). The earliest long-distance obsidian transport: Evidence from the ~200 ka Middle Stone Age Sibilo School Road site, Baringo, Kenya. *Journal of Human Evolution, 103*, 1–19.

Boëda, E. (1995). Levallois: A volumetric construction, methods, a technique. In H. Dibble & O. Bar-Yosef (Eds.), *The definition and interpretation of levallois technology* (pp. 41–68). Madison: Prehistory Press.

Bofill, M. (2012). Quantitative analysis of use-wear patterns: A functional approach to the study of grinding stones. *Broadening Horizons, 3*, 57–78.

Bordes, F. (1961). *Typologie du paléolithique ancien et moyen*. Bordeaux: Imprimeries Delmas.

Borel, A., Ollé, A., Vergès, J. M., & Sala, R. (2014). Scanning Electron and optical light microscopy: Two complementary approaches for the understanding and interpretation of usewear and residues on stone tools. *Journal of Archaeological Science, 48*, 46–59.

Borradaile, G. J., Kissin, S. A., Stewart, J. D., Ross, W. A., & Werner, T. (1993). Magnetic and optical methods for dating the heat treatment of chert. *Journal of Archaeological Science, 20*, 57–66.

Bostwick, T. W., & Burton, J. H. (1993). A study in sourcing Hohokam basalt ground stone implements. *Kiva, 58*, 357–372.

Brandl, M., Martinez, M. M., Hauzenberger, C., Fitzmoser, P., Nymoen, P., & Mehler, N. (2018). A multi-technique analytical approach to sourcing Scandinavian flint: Provenance of ballast flint from the shipwreck 'Leirvigen 1', Norway. *PLoS ONE, 13*(8), e0200647.

Braswell, G. E., Clark, J. E., Aoyama, K., McKillop, H. I., & Glascock, M. D. (2000). Determining the provenance of obsidian artifacts from the Maya region: A test of the efficacy of visual sourcing. *Latin American Antiquity, 11*, 269–282.

Bretzke, K., & Conard, N. J. (2012). Evaluating morphological variability in lithic assemblages using 3D models of stone artifacts. *Journal of Archaeological Science, 39*, 3740–3749.

Burton, J. (1980). Making sense of waste flakes. *Journal of Archaeological Science, 7*, 131–148.

Carballo, D. M., Carballo, J., & Neff, H. (2007). Formative and Classic period obsidian procurement in central Mexico: A compositional study using laser ablation-inductively couled plasma-mass spectrometry. *Latin American Antiquity, 18*(1), 27–43.

Carelli, P., & Kresten, P. (1997). Give us this day our daily bread: A study of late Viking Age and Medieval quernstones in south Scandinavia. *Acta Archaeologica, 68*, 109–137.

Chazan, M. (1997). Redefining Levallois. *Journal of Human Evolution, 33*, 709–735.

Chazan, M. (2009). Pattern and technology: Why the chaîne opératoire matters. In J. J. Shea & D. E. Lieberman (Eds.), *Transitions in prehistory: Essays in honor of Ofer Bar-Yosef* (pp. 467–498). Oxford: Oxbow Books.

Clark, J. D. (1970). *The prehistory of Africa*. London: Thames & Hudson.

Close, A. E. (1978). The identification of style in lithic artifacts. *World Archaeology, 10*, 223–236.

Cotterell, B., & Kamminga, J. (1990). *Mechanics of pre-industrial technology*. Cambridge: Cambridge University Press.

Crabtree, D. E. (1968). Mesoamerican polyhedral cores and prismatic blades. *American Antiquity, 33*(4), 446–478.

Davis, W. (1990). Style and history in art history. In M. Conkey & C. A. Hastorf (Eds.), *The uses of style in archaeology* (pp. 18–31). Cambridge: Cambridge University Press.

Deschamps, M., & Zilhão, J. (2018). Assessing site formation and assemblage integrity through stone tool refitting at Gruta da Oliveira (Almonda karst system, Torres Novas, Portugal): A Middle Paleolithic case study. *PLoS ONE, 13*(2), e0192423.

Dibble, H. L. (1985). Technological aspects of flake variation. *American archaeology, 5*, 236–240.

Dibble, H. L. (1995). Middle Paleolithic scraper reduction: Background, clarification, and review of evidence to date. *Journal of Archaeological Method and Theory, 2*(4), 299–368.

Dibble, H. L., & Bernard, M. C. (1980). A comparative study of basic edge angle measurement techniques. *American Antiquity, 45*(4), 857–865.

Dickson, F. P. (1976). Australian ground stone hatchets: Their design and dynamics. *Australian Archaeology, 5*, 33–48.

Dickson, F. P. (1980). Making ground stone tools. *Archaeology and Physical Anthropology in Oceania, 15*(3), 162–167.

Dixon, J. E., Cann, J. R., & Renfrew, C. (1968). Obsidian and the origins of trade. *Scientific American, 218*(3), 38–46.

Dogandzic, T., Braun, D. R., & McPherron, S. P. (2015). Edge length and surface area of a blank: Experimental assessment of measures, size predictions and utility. *PLoS ONE, 10*(9), e0133984.

Domanski, M., & Webb, J. (2007). A review of heat treatment research. *Lithic Technology, 32*(2), 153–194.

Dubreuil, L., & Savage, D. (2014). Ground stones: A synthesis of the use-wear approach. *Journal of Archaeological Science, 48*, 139–153.

Evans, A. A. (2014). On the importance of blind testing in archaeological science: The example from lithic functional studies. *Journal of Archaeological Science, 48*, 5–14.

Evans, A. A., & Donahue, R. E. (2008). Laser scanning confocal microscopy: A potential technique for the study of lithic microwear. *Journal of Archaeological Science, 35*, 2223–2230.

Faulks, N., Kimball, L. R., Hidjrati, N., & Coffey, T. (2011). Atomic force microscopy of microwear traces on Mousterian tools from Jyshtylagty Lagat (Weasel Cave), Russia. *Scanning, 33*, 304–315.

Fischer, A., Vemming, P., & Rasmussen, P. (1984). Macro and micro wear traces on lithic projectile points. *Journal of Danish Archaeology, 3*, 19–46.

Flenniken, J. J., & White, J. P. (1983). Heat treatment of silicious rocks and its implications for Australian prehistory. *Australian Aboriginal Studies, 1*, 43–47.

Gabriel, I., & Kempke, T. (2011). *Starigard/Oldenburg, Hauptburg der Slawen in Wagrien VI. Die Grabfunde. Einführung und archäologisches Material*. Neumünster: Offa-Bücher 85.

Geneste, J. M., David, B., Plisson, H., Delannoy, J. J., & Petchey, F. (2012). The origins of ground-edge axes: New findings from Nawaria Gabarnmang, Arnhem Land (Australia) and global implictions for the evolution of fully modern humans. *Cambridge Archaeological Journal, 22*(1), 1–17.

Glascock, M. D. (Ed.). (2002). *Geochemical evidence for long-distance exchange*. Westport: Bergin and Garvey.

Golitko, M., Meiehoff, J., & Terrell, J. E. (2010). Chemical characterization of obsidian from the Sepik Coast (PNG). *Archaeology in Oceania, 45*(3), 120–129.

Goodale, N., & Andrefsky, W. (Eds.). (2015). *Lithic technological systems: Stone, human behavior, evolution*. Cambridge: Cambridge University Press.

Grace, R. (1989). *Interpreting the function of stone tools: The quantification and computerization of microwear analysis*. Oxford: British Archaeological Reports.

Grace, R. (1996). Use-wear analysis: The state of the art. *Archaeometry, 38*, 209–229.

Gwinnet, A. J., & Gorelick, L. (1987). The change from stone drills to copper drills in Mesopotamia. *Expedition, 29*, 15–24.

Hill, J., & Gunn, J. (Eds.). (1977). *The individual in prehistory*. New York: Academic.

Hiscock, P. (2004). Slippery and Billy: Intention, selection and equifinality in lithic artefacts. *Cambridge Archaeological Journal, 14*(1), 71–77.

Hiscock, P., & Tabrett, A. (2010). Generalization, inference and the quantification of lithic reduction. *World Archaeology, 42*(4), 545–561.

Hodder, I. (1982). *Symbols in action: Ethnoarchaeological studies of material culture*. Cambridge: Cambridge University Press.

Hodges, H. (1964). *Artifacts: An introduction to early materials and technology*. London: J. Baker.

Horsfall, G. A. (1987). Design theory and grinding stones. In B. Hayden (Ed.), *Lithic studies among the contemporary highland Maya* (pp. 332–377). Tucson: University of Arizona Press.

Hranicky, W. J. (2013). *Archaeological concepts, techniques, and terminology for American -prehistoric lithic technology*. Bloomington: AuthorHouse.

Hruby, Z. X. (2007). Ritualized chipped-stone production at Piedras Negras, Guatemala. In Z. X. Hruby & R. K. Flad (Eds.), *Rethinking craft specialization in complex societies: Arcaheological analyses of the social meaning of production* (pp. 68–87). Archaeological Papers of the American Anthropological Association 17. Washington, DC: American Anthropological Association.

Huckell, B. (1986). *A ground stone implement quarry on the lower colorado river, Northwestern Arizona*. Phoenix: Bureau of Land Management.

Hughes, R. E. (1994). Intrasource chemical variability of artefact-quality obsidians from the Casa Diablo area, California. *Journal of Archaeological Science, 21*(2), 263–271.

Hughes, S. S. (1998). Getting to the point: Evolutionary change in prehistoric weaponry. *Journal of Archaeological Method and Theory, 5*, 345–408.

Hutchings, W. K. (1991). *The Nachcharini composite projectile point: Design theory and the study of hunting systems technology at Mugharet-Nachcharini in the post-Natufian Levant*. M. A. thesis, University of Toronto.

Hyland, D. C., Tersak, J. M., Adovasio, J. M., & Siegel, M. I. (1990). Identification of the species of origin of residual blood on lithic material. *American Antiquity, 55*(1), 104–112.

Inizan, M. L., Roche, H., & Tixier, J. (1992). *Technology of knapped stone*. Meudon.

Jackson, S. E. (2017). Envisioning artifacts: A Classic Maya view of the archaeological record. *Journal of Archaeological Method and Theory, 24*, 579–610.

Jeter, M. D. (1980). Analysis of flaked stone artifacts and debitage. In D. E. Doyel & S. S. Debowski (Eds.), *Prehistory in Dead Valley, East-central Arizona: The TG&E Springerville Report* (pp. 235–304). Tucson: Arizona State Museum.

Jezek, M. (2013). Touchstones of archaeology. *Journal of Anthropological Archaeology, 32*, 713–731.

Kadowaki, S., & Banning, E. B. (2018). Morphometric and refitting analyses of flaked stone artifacts from Tabaqat al-Bûma and al-Basatîn, northern Jordan: Sickle elements and core-reduction technology in the late Neolithic (6th millennium BCE) in the southern Levant. *Journal of Archaeological Science: Reports, 19*, 64–79.

Kingery, W. D. (2001). The design process as a critical component of the anthropology of technology. In M. B. Schiffer (Ed.), *Anthropological perspectives on technology* (pp. 123–138). Albuquerque: University of New Mexico Press.

Knecht, H. (Ed.). (1997). *Projectile technology*. New York: Plenum Publishing.

Kooyman, B., Newman, M. E., & Ceri, H. (1992). Verifying the reliability of blood residue analysis on archaeological tools. *Journal of Archaeological Science, 19*(3), 265–269.

Kukan, G. J. (1978). *A technological and stylistic study of microliths from certain epipaleolithic assemblages*. PhD thesis, University of Toronto.

Kuzmin, Y. V., Glascock, M. D., & Izuho, M. (2013). The geochemistry and major sources of archaeological obsidian of Hokkaido Island (Japan): Shirataki and Oketo. *Archaeometry, 55*(3), 355–369.

Latorre, A. M., Pérez, A. P., Gibaja Bao, J. F., Zamora, G. R., & Gómez-Gras, D. (2017). Use-wear analysis of Neolithic polished axes and adzes: the site of "Bòbila Madurell-Can Gambús-1-2" (northeast Iberian peninsula). *Quaternary International, 427*, 158–174.

Lechtman, H. (1977). Style in technology: Some early thoughts. In H. Lechtman & R. Merrill (Eds.), *Material culture: Styles, organization, and dynamics of technology* (pp. 3–20). St. Paul: Proceedings of the American Ethnological Society.

Lemonnier, P. (1986). The study of material culture today: Toward an Anthropology of technical systems. *Journal of Anthropological Archaeology, 5*, 147–186.

Leroi-Gourhan, A. (1964). *Le Geste et la Parole I: Technique et Langage*. Paris: Albin Michel.

Lewis, R., Tsoraki, C., Broughton, J., Cripps, J. C., Afodun, S. A., Slatter, T., et al. (2011). Abrasive and impact wear of stone used to manufacture axes in Neolithic Greece. *Wear, 271*, 2549–2560.

Lin, S. C., Douglass, M. J., Holdaway, S., & Floyd, B. (2010). The application of 3D laser scanning technology to the assessment of

ordinal and mechanical cortex quantification in lithic analysis. *Journal of Archaeological Science, 37*, 694–702.
Lindner, C. C., & Rodger, C. A. (2009). *Design theory* (2nd ed.). Boca Raton: CRC Press.
Lowenstein, J. M. (1985). Molecular approaches to the identification of species. *American Scientist, 73*, 541–547.
Luedtke, B. E. (1978). Chert sources and trace element analysis. *American Antiquity, 43*, 413–423.
Luedtke, B. E. (1979). The identifications of sources of chert artifacts. *American Antiquity, 44*, 744–757.
Lyman, R. L., & VanPool, T. (2009). Metric data in archaeology: A study of intra-analyst and inter-analyst variation. *American Antiquity, 74*(3), 485–504.
Macdonald, D. A. (2014). The application of focus variation microscopy for lithic use-wear Quantification. *Journal of Archaeological Science, 48*, 26–33.
Maget, M. (1962). *Guide d'Étude directe de comportements culturels* (2nd ed.). Paris: Centre national de la Recherche Scientifique.
Mahaney, R. A. (2014). Lithic analysis as a cognitive science: A framework. *Lithic Technology, 39*(3), 73–189.
Marreiros, J. M., Gibaja Bao, J. F., & Fereira Bicho, N. (Eds.). (2015). *Use-wear and residue analysis in archaeology*. New York: Springer.
Miyamoto, A., & Murakami, Y. (2000). The morphology of Hertzian cone in plate glass by projectile impact. *The Society of Materials Science Japan, 49*(8), 867–876.
Moreau, L., Brandl, M., Filzmoser, P., Hauzenberger, C., Goemaere, É., Jadin, I., et al. (2016). Geochemical sourcing of flint artifacts from western Belgium and the German Rhineland: Testing hypotheses on Gravettian period mobility and raw material economy. *Geoarchaeology, 31*, 229–243.
Morgan, T. J. H., Uomini, N. T., Rendell, L. E., Chouinard-Thuly, L., Street, S. E., Lewis, H. M. C., et al. (2014). Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nature Communications*. https://doi.org/10.1038/ncomms7029.
Morrow, T. A. (1997). A chip off the old block: Alternative approaches to debitage analysis. *Lithic Technology, 22*(1), 1–69.
Muller, A., & Clarkson, C. (2014). Estimating original flake mass on blades using 3D platform area: Problems and prospects. *Journal of Archaeological Science, 5*, 31–38.
Muller, A., & Clarkson, C. (2018). A new method for accurately and precisely measuring flake platform area. *Journal of Archaeological Science, Reports, 8*, 178–186.
Nadel, D. (1993). Patination of flint artefacts: Evidence from Bikta, a submerged prehistoric occurrence in the Sea of Galilee, Israel. *Journal of the Israel Prehistoric Society, 25*, 145–162.
Nelson, M. C. (1987). Site content and structure: Quarries and workshops in the Maya Highlands. In B. Hayden (Ed.), *Lithic studies among the contemporary highland Maya* (pp. 120–147). Tucson: University of Arizona Press.
Nelson, M. C. (1991). The study of technological organization. *Archaeological Method and Theory, 3*, 57–100.
Odell, G. H. (2012). *Lithic analysis*. New York: Springer.
Pavlish, L. A. (1978). Thermoluminscent (TL) determination of prehistoric heat treatment of chert artefacts. *Science, 197*, 1359–1362.
Pelegrin, J. (1993). A framework for analysing prehistoric stone tool manufacture and a tentative application to some early stone industries. In A. Berthelet & J. Cavaillon (Eds.), *The use of tools by human and non-human primates* (pp. 302–314). Oxford: Clarendon Press.
Phillips, C. S., & Simpson, St. J. (2018). *Softstone: Approaches to the study of chlorite and calcite vessels in the Middle East and Central Asia from prehistory to the present* (British Foundation for the study of Arabia monographs 20). Oxford: Archaeopress.
Pye, D. (1964). *The nature of design*. London: Studio Vista.
Reimer, R. (2015). Reassessing the role of Mount Edziza obsidian in northwestern North America. *Journal of Archaeological Science, Reports, 2*, 418–426.
Robins, G. V., Seeley, N. J., McNeil, D. A. C., & Symons, M. R. C. (1978). Identification of ancient heat treatment in flint artefacts by ESR spectroscopy. *Nature, 276*, 703–704.
Rosenberg, D., Rowan, Y., & Gluhak, T. (2016). Leave no stone unturned. Perspectives on ground stone artifact research. *Journal of Lithic Studies, 3*(3), 1–15.
Rowan, Y. M., & Ebeling, J. R. (Eds.). (2008). *New approaches to old stones: Recent studies of ground stone artifacts*. London: Equinox.
Rutter, G. P., Pearson, D. G., Philip, G., Day, J. M. D., & Ottley, C. J. (2003). The use of ICP-MS in provenancing igneous stone artefacts: Examples from the southern Levant. In G. Holland & S. D. Tanner (Eds.), *Plasma source mass spectrometry: Applications and emerging technologies* (pp. 209–220). Cambridge: The Royal Society of Chemistry.
Sackett, J. R. (1977). The meaning of style in archaeology. *American Antiquity, 42*, 369–380.
Sackett, J. R. (1982). Approaches to style in lithic archaeology. *Journal of Anthropological Archaeology, 1*, 59–112.
Sackett, J. R. (1990). Style and ethnicity in archaeology: The case for isochrestism. In M. Conkey & C. A. Hastorf (Eds.), *The uses of style in archaeology* (pp. 32–43). Cambridge: Cambridge University Press.
Sánchez de la Torre, M., Gratuze, B., Mangado, X., & Delage, C. (2019). The lithic landscape around Kharaneh IV (Azraq Basin, Jordan): Petrographical and geochemical characterization of geological cherts. *Journal of Archaeological Science, Reports, 26*. https://doi.org/10.1016/j.jasrep.2019.05.022.
Sax, M., & Meeks, N. D. (1995). Methods of engraving Mesopotamian quartz cylinder seals. *Archaeometry, 37*, 25–36.
Sax, M., Meeks, N. D., & Collon, D. (2000). The early development of the lapidary engraving wheel in Mesopotamia. *Iraq, 62*, 157–176.
Sax, M., Meeks, N. D., Michaelson, C., & Middleton, A. P. (2004). The identification of carving techniques on Chinese jade. *Journal of Archaeological Science, 31*, 1413–1428.
Schiffer, M. B. (1976). *Behavioral archaeology*. New York: Academic Press.
Schlanger, N. (2004). The *chaîne opératoire*. In C. Renfrew & P. Bahn (Eds.), *Archaeology: The key concepts* (pp. 25–31). Boca Raton: CRC Press.
Schmidt, P., Léa, V., Sciau, P., & Frölich, F. (2013). Detecting and quantifying heat treatment of flint and other silica rocks: A new non-destructive method applied to heat-treated flint from the Neolithic Chassey culture, southern France. *Archaeometry, 55*(5), 794–805.
Schneider, J. S., Lerch, M. K., & Smith, G. A. (1996). Quarrying and production of milling implements at Antelope Hill, Arizona. *Journal of Field Archaeology, 23*, 299–310.
Sellet, F. (1993). Chaîne opératoire: The concept and its applications. *Lithic Technology, 18*, 106–112.
Semenov, S. A. (1964). *Prehistoric technology: An experimental study of the oldest tools and artefacts from traces of manufacture and wear* (M. W. Thompson, Trans.). Bradford-on-Avon: Moonraker Press.
Shackley, M. S., & Sahle, Y. (2017). Geochemical characterization of four Quaternary obsidian sources and provenance of obsidian artifacts from the Middle Stone Age site of Gademotta, Main Ethiopian Rift. *Geoarchaeology, 32*(2), 302–310.
Shea, J. J. (2013). *Stone tools in the Paleolithic and Neolithic Near East: A guide*. Cambridge: Cambridge University Press.
Shea, J., & Klenck, J. (1993). An experimental investigation of the effects of trampling on the results of lithic microwear analysis. *Journal of Archaeological Science, 20*(2), 175–194.

Shennan, S. (2002). *Genes, memes and human history*. London: Thames & Hudson.

Shennan, S. (2009). Evolution in archaeology. *Annual Review of Anthropology, 37*(1), 75–91.

Sheppard, P. J. (1987). *The Capsian of North Africa. Stylistic variation in stone tool assemblages*. Oxford: British Archaeological Reports.

Shott, M. J. (2003). *Chaîne opératoire* and reduction sequence. *Lithic Technology, 28*(2), 95–105.

Skibo, J. M., & Schiffer, M. B. (2008). *People and things: A behavioral approach to material culture*. New York: Springer.

Soressi, M., & Geneste, J.-M. (2011). The history and efficacy of the *chaîne opératoire* approach to lithic analysis: Studying techniques to reveal past societies in an evolutionary perspective. *PaleoAnthropology, 2011*, 334–350. https://doi.org/10.4207/PA.2011.ART63.

Speer, C. A. (2014). LA-ICP-MS analysis of Clovis period projectile points from the Gault site. *Journal of Archaeological Science, 52*, 1–11.

Speth, J. D. (1972). The mechanical basis of percussion flaking. *American Antiquity, 37*(1), 34–60.

Stahle, D., & Dunn, J. (1982). An analysis and application of the size distribution of waste flakes from the manufacture of bifacial stone tools. *World Archaeology, 14*, 84–97.

Stemp, W. J., Childs, B. E., Vionnet, S., & Brown, C. A. (2009). Quantification and discrimination of lithic use-wear: Surface profile measurements and length-scale fractal analysis. *Archaeometry, 51* (3), 366–382.

Stemp, W. J., Watson, A., & Evans, A. A. (2016). Topical review: Surface analysis of stone and bone tools. *Surface Topography: Metrology and Properties, 4*(1), 1–25. http://iopscience.iop.org/article/10.1088/2051-672X/4/1/013001

Stemp, W. J., Peuramaki-Brown, M., & Awe, J. J. (2019). Ritual economy and ancient Maya bloodletting: Obsidian blades from Actun Uayazba Kab (Handprint Cave), Belize. *Journal of Anthropological Archaeology, 53*, 304–324.

Stocks, D. (2003). *Experiments in Egyptian archaeology: Stoneworking technology in ancient Egypt*. London: Routledge.

Sullivan, A. P., & Rozen, K. C. (1985). Debitage analysis and archaeological interpretation. *American Antiquity, 50*(4), 755–779.

Takashi, T. (2012). MIS3 edge-ground axes and the arrival of the first *Homo sapiens* in the Japanese archipelago. *Quaternary International, 248*, 70–78.

Tixier, J. (2012). *Méthode pour l'Étude des outillages lithiques*. Luxembourg: Centre National de Recherche Archéologique du Luxembourg.

Van Peer, P. (1992). *The Levallois reduction strategy*. Madison: Prehistory Press.

Vaughan, P. (1985). *Use-wear analysis of flaked stone tools*. Tucson: University of Arizona Press.

Vierra, B. J. (1995). *Subsistence and stone tool technology: An Old World perspective*. Madison: Prehistory Press.

Wagalawatta, T., Bebermeier, W., Kohlmeyer, K., & Schütt, B. (2016). An ethnoarchaeological study of stone quarrying techniques in historical Anuradhapura. *Ethnoarchaeology, 8*(2), 158–179.

Watson, A., & Nash, D. J. (1997). Desert crusts and varnishes. In D. S. G. Thomas (Ed.), *Arid zone geomorphology: Process, form, and change in drylands* (pp. 69–107). New York: John Wiley & Sons.

Weisler, M. I., Bolhar, R., Ma, J. L., St. Pierre, E. J., Sheppard, P. J., Walter, R., et al. (2016). Cook Island artifact geochemistry demonstrates spatial and temporal extent of pre-European interarchipelago voyaging in East Polynesia. *PNAS, 113*(29). https://doi.org/10.1073/pnas.1608130113.

White, R. (1993). Introduction to *Gesture and Speech* by A. Leroi-Gourhan, trans. A. B. Berger. Cambridge: MIT Press.

Whittaker, J. C. (1994). *Flintknapping: Making and understanding stone tools*. Austin: University of Texas Press.

Wiessner, P. (1983). Style and social information in Kalahari San projectile points. *American Antiquity, 49*, 253–276.

Wiessner, P. (1985). Style or isochrestic variation? A reply to Sackett. *American Antiquity, 50*, 160–166.

Wiessner, P. (1990). Is there a unity to style? In M. Conkey & C. A. Hastorf (Eds.), *The uses of style in archaeology* (pp. 105–112). Cambridge: Cambridge University Press.

Wobst, M. (1977). Stylistic behavior and information exchange. *Anthropological Papers of the University of Michigan, 61*, 317–342.

Wright, K. A. (1992). A classification system for ground stone tools from the prehistoric Levant. *Paléorient, 18*, 53–81.

Wright, K. A. (2014). Domestication and inequality? Households, corporate groups and food processing tools at Neolithic Çatalhöyük. *Journal of Anthropological Archaeology, 33*, 1–33.

12 Ceramic Artifacts

Potsherds have been called the alphabet of the archaeologist
Bibby (1969: 50).

"Pottery" in an archaeological context most often concerns the fragments, or sherds, of ceramic vessels, but archaeological ceramics also include such things as roof tiles, drain tiles, coffins, ossuaries, stands, lamps and figurines, many of which are vessels only in the broadest sense of the term. This chapter will emphasize ceramic containers. Ceramics tend to be ubiquitous on sites of the mid- to late Holocene in many parts of the world, preserve almost as well as lithics, and provide a great deal of latitude for stylistic variation, making them very important sources of information. For archaeologists, "ceramics" include earthenwares, terracottas, stonewares, porcelains, and other materials made from fired clay, while materials scientists use the term differently (Kingery et al. 1976; Rice 2015: 3–4). Archaeologists have tended to analyze the interrelated technological, functional, and stylistic dimensions of ceramics and their associations with chronological, spatial, social, economic and ideological differences among the makers of these artifacts.

Technological analyses of ceramics focus on the selection, acquisition and preparation of raw materials, the manufacturing techniques and sequences, modification or recycling of finished artifacts, and the artifacts' eventual failure or loss and post-depositional history (Hunt 2017; Montana 2017; Roux 2017; Waksman 2017). Unlike lithic technology, pottery technology is not primarily a reductive one, although some reduction can occur, in thinning vessel walls, for example. Technological variation can be related to differences in the cost or availability of materials, to the skills, knowledge, preferences, motor habits, and traditions of potters, to the anticipated uses of finished products, and to their anticipated use-life.

As with other artifact types, functional variability results from design considerations that affect the usefulness of artifacts for various tasks, such as containment, temperature control, transport, distribution and consumption, social display, and disposal, as well as their actual use in tasks for which they may or may not have been designed. Some kinds of pottery show considerable investments in design, but pottery is also versatile and can be used for many tasks for which their designers did not intend them. A Greek amphora designed to contain and transport wine, for example, could be recycled to ship fish paste or even be used as building material in a wall or roof, while sherds from broken pots can be made into spindle whorls or scraping tools. Design criteria for pottery may include aspects of the shape and size of a vessel or its parts, but also characteristics of its material, such as the porosity or strength of vessel walls.

12.1 Anatomy of a Pot

The segmentation and orientation rules (see p. 61) for pottery and similar non-ceramic vessels (e.g., glass bottles) have developed over many decades and require unambiguous definition. Exact definitions will vary to some extent with researchers' questions and the idiosyncrasies of the assemblages they study, but most archaeologists share at least broad concepts for the parts of a vessel.

Orientation rules typically depend on the likely orientation of the vessel when it was in use, typically with the orifice or opening at the top (and most often horizontal), and the base, which would normally be in contact with a surface on which the vessel rested, at the bottom. Archaeologists call this orientation "at stance." However, things are not always so simple, as some pots have unusual or asymmetrical shapes or orifices that are not horizontal and possibly not even at the top. Others have multiple orifices or are tube-like with no obvious base and ambiguity about which orifice should be

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_12

up. Some that look like bowls may have been used as lids that are properly oriented with the rim at the bottom. Still others, such as drains or smoking pipes, were typically oriented horizontally or nearly so while in use, whether or not archaeologists want to represent them in that way. Things become even more difficult when we are dealing, as we usually are, with vessel fragments rather than whole vessels; unless it is attached to a rim or base, a loop handle could have had vertical, horizontal, or even diagonal orientation. Finally, a pot's orientation "in use" is actually variable; a cooking pot, for example, would have its orifice upward during cooking, but downward while drying or being stored, and even tilted 45° for roasting beans (Skibo 1992: 64–73).

Archaeologists adopt conventions for orienting ceramics and their fragments that do not always match the way vessels would have been oriented during use. A typical set of orientation rules would be as follows:

- Rim sherds (except for lids) are oriented with the rim upwards and horizontal ("at stance"). In addition, we usually assume that the orifice is circular in plan (transverse) view.
- Base sherds (except round or pointed bases) are oriented downward in such a way as to maximize contact between the base and the (assumed) surface on which it rests ("at stance").
- Pointed bases are oriented point-down and with their long axis vertical (except for tripod bases, whose legs may angle outwards from the vessel center).
- Round bases are oriented so that the thickest part of the base touches the (assumed) surface on which the base would lie, even though they would not normally be used on flat surfaces.
- Body sherds, where possible, are oriented so that signs of coil construction, smoothing, rouletting, or wheel throwing are predominantly horizontal. This often leaves it undetermined which way is "up." In that event, one common convention is to put the thickest part of the sherd downward.
- Handles, spouts, or knobs that are not attached to a rim or base that provides evidence for orientation are oriented in the direction that is most common among relatively complete vessels in the same assemblage (this has the potential of causing bias, however).

For most measurements of rims and bases, the most important element of orientation is the stance. To stance a rim sherd, hold it upside-down against a flat surface and rock it back and forth until you find the position that puts all or most of the lip in contact with the surface. The flat surface is then assumed to represent the horizontal plane of the pot (see Fig. 21.13 for how to draw a sherd at stance).

To describe slices through vessels or their sherds, we can use the same terms as for wood, charcoal, or glass bottles. The **radial plane** is any slice that radiates from the vertical (central) axis of the pot's radial or bilateral symmetry (Fig. 12.1), while the **transverse plane** is any slice along a plane parallel to the stance (i.e., horizontal if the pot is sitting on a flat surface), and the lateral or **tangential view** is any view that is tangent to the surface of the pot, and a **tangential plane** is parallel to that tangent.

Segmentation rules for pottery vary internationally but tend toward the following (see also Rice 2015: 232–244):

- Lip: The narrow surface most distant from the base as measured along the center of the vessel walls. The lip

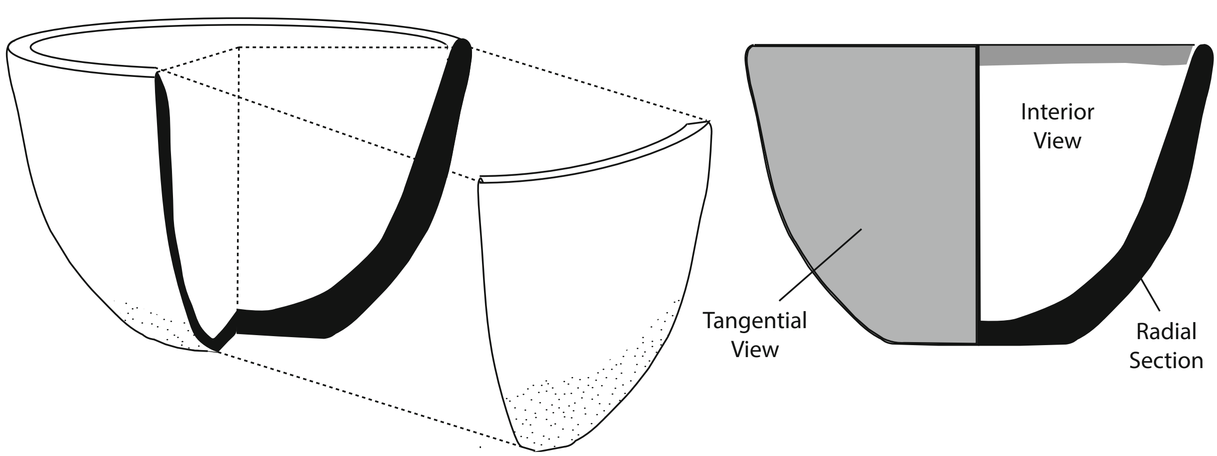


Fig. 12.1 Archaeological descriptions of pottery emphasize the radial section (a cross-section parallel to the pot's axis of symmetry to provide a profile from rim to base) and tangential view (i.e., view of the vessel's exterior surface). A transverse view (looking down at the rim or up at the base) is more important for vessels that are not circular in plan or that have unusual features on base, interior, or rim. Illustrations of pottery typically employ "cutaway" views that show a radial section as well as a reconstruction of the exterior and interior surfaces of the vessel, as though the illustrator had cut away one-quarter of the vessel (left)

would be mostly in contact with the plane used to stance a rim.

- Rim: the portion of the vessel closest to the vessel orifice (including the lip).
- Neck: A constriction of the vessel below the orifice but above the maximum diameter of the body or shoulder.
- Shoulder: The region above the body's maximum diameter but below the neck or rim.
- Handle: An appendage attached to the (usually exterior) wall of the body, neck or rim that apparently aided in the vessel's manipulation. Often there are two or more handles on a vessel.
- Base: the portion from the lower part of the body to the points that would normally be in contact with a surface when the vessel is at rest (e.g.,—, floor, stand, or tripod).
- Foot: An appendage attached along the circumference of the base that raises the body.
- Tripod: A set of three knob-like or leg-like supports attached along the circumference of the base that raises the body.
- Spout: Neck-like appendage to restrict a secondary orifice that occurs to one side of the vessel's vertical axis (typically on a vessel's upper body or shoulder).

Where pottery exhibits decoration, it is also desirable to have orientation and segmentation rules for the decoration (see Figs. 4.3 and 12.2).

12.2 Measuring Vessel Form and Size

Archaeologists routinely make many observations that are relevant to vessels' function, use or style, as well as their manufacture (e.g., Sinopoli 1991: 56–65). Many of these concern size and morphology (Table 12.1).

Some of the attributes that archaeologists measure pertain to whole vessels, when those occur. A useful one is the ratio of orifice diameter to vessel height. Vessels with large orifices relative to their height are called "open" vessels, while ones with relatively small orifices are called "closed" or "restricted" vessels (Fig. 12.3). Although other general shape measures are available, this one has the advantage of being related to the relative accessibility of vessel contents, something closely related to the vessel's designed function. In a wine amphora, for example, its closed or restricted form (generally with a narrow neck and small orifice) helps to contain the wine, allow sealing, reduce spillage, and facilitate pouring. Cooking pots, by contrast, need to be somewhat

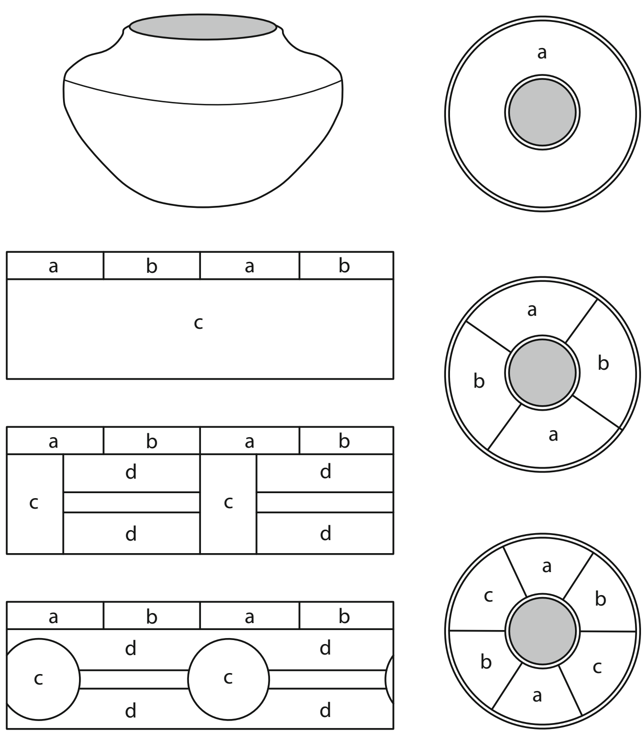


Fig. 12.2 An example of segmentation rules for painted decoration on the exterior of Pueblo vessels (after Bunzel 1972). Vessels like the one at upper left are divided into zones of actual or potential decoration both vertically and radially. The rectangles show the segmentation of decorative panels as viewed tangentially (from the side). The circles at right show the radial segments as viewed transversely (from the top)

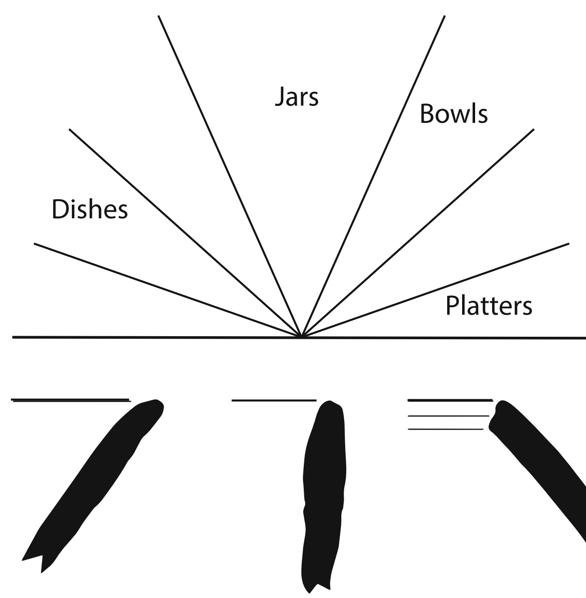


Fig. 12.3 A simple template (top) for classifying vessels by the degree of "openness" for whole and reconstructable vessels (after Orton 1980: 34). The radii originate at the center of the vessel's bottom and extend to the top of the rim of the orifice. However, for rim sherds we can only classify openness with reference to the upper vessel (bottom), with everted, vertical, and inverted openings

Table 12.1 Examples of some common ceramic formal attributes and their measurement

Attribute	Scale	Selected purposes
Internal rim diameter	Ratio	Size, function, EVE, MNV
Preserved circumference	Ratio	Preservation, EVE, MNV
Neck diameter	Ratio	Size, function
Neck height	Ratio	Function
Maximum body diameter	Ratio	Size, function
Rim thickness	Ratio	Style
Rim shape	Nominal	Style, function
Rim angle	Ratio or ordinal	Function, style
Body wall thickness	Ratio	Size, function
Handles	Dichotomous or nominal	Function, style
Base type	Nominal	Function, style
Base diameter	Ratio	Size, function

Modified from Sinopoli (1991: 61), along with selected purposes, including some quantitative measures of abundance (see Chap. 7). Other attributes of interest are non-formal, such as mineral characteristics of the paste and evidence for firing conditions

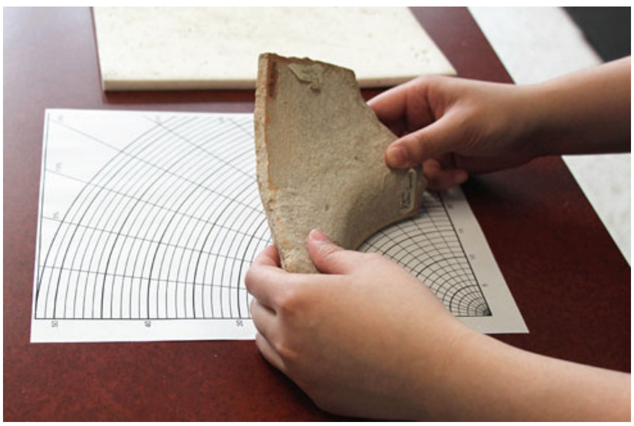


Fig. 12.4 Use of a diameter chart to estimate the diameters of circular rims or bases as well as their maximum horizontal preservation as a percentage of a full circle

more open to allow stirring as well as addition and removal of ingredients and meals. Vessels for serving and consuming food are typically open (bowls and plates).

To measure rim or base diameters, archaeologists often use a diameter chart (Fig. 12.4). This has nested arcs corresponding with various diameters, typically at 1 cm intervals (or 0.5 cm of radius), as well as radial line segments to indicate the proportion of the circumference that each rim or base preserves, a useful measure for calculating EVE (see Chap. 7) or degree of fragmentation. To use it, you would "stance" a sherd so that the entire rim or base makes contact with the chart and move it around until it matches one of the arcs while one end of the sherd touches the chart's "y-axis."

Fortunately for archaeologists, most archaeological vessels had almost circular orifices, so that the method just described is sometimes imprecise, but reasonable accurate. However, vessels with oval or other non-circular orifices did exist in some cultures. In cases where oval vessels are possible or likely, our estimates of vessel size based on rim diameter from sherds could be, not only imprecise, but also biased, because substantial overestimates of diameter on the longer sides of the oval would outweigh the smaller underestimates of diameter on the ends. For rectangular vessels, meanwhile, we have little or no basis for estimating vessel size except when the vessels are nearly complete.

Increasingly, archaeologists are able to make 3D scans of sherds that allow much more complex measurement of shape (Fig. 12.5), much of it automated (Di Angelo et al. 2018; Karaskik and Smilansky 2008; Martínez-Carrillo and Barceló 2017; Wilczek et al. 2018).

12.3 Ceramic Ecology and *Chaînes Opératoires* of Ceramic Vessels

Matson (1965) pioneered a concept he called "ceramic ecology" as a framework for studying archaeological pottery in a way that somewhat resembles *chaînes opératoires* (see pp. 214–215) except for its greater focus on the ecological environment of pottery production, perhaps over-emphasis of technological constraints (Gosselain 1998), and less emphasis (but still some) on gestures, skills and knowledge. This approach to studying archaeological pottery is structured around the selection, acquisition, and processing of materials, the forming, finishing, and decoration of vessels, and the drying, firing, sometimes re-firing, cooling, and sometimes sealing of vessels to make them less porous. The ecological aspect has a strong role in both the sourcing of raw materials and the timing of pottery production, which depends a great deal on weather and seasonality in both climate and human activities, such as agricultural tasks, that compete for potters' time. *Chaînes opératoires* place less emphasis on the environmental variables but share the focus on techniques and sequences of technological processes, while also including the disuse, discard and recycling of pottery, although many

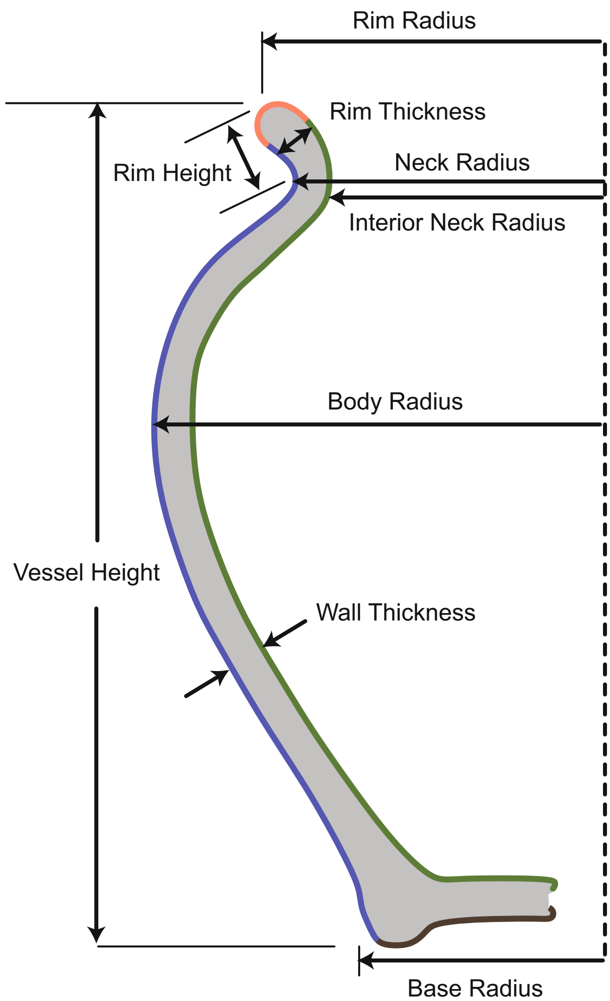


Fig. 12.5 Examples of measurements of rim shape and various wall thicknesses and diameters on a radial profile from a 3D scan of a vessel that has been virtually "sectioned". (Y. Salama, after Di Angelo et al. 2018)

researchers omit this last point in favor of concentrating on the steps up to a finished vessel (Roux 2017). Techniques involve choosing the source of energy (e.g., squeezing with fingers or rotation on a wheel), discontinuous or continuous pressure, and assembled parts or a single mass of clay body (Roux and Courty 1998: 763). While the descriptions below may seem very linear, in fact a *chaîne opératoire* can involve repeating some steps, carrying out some steps simultaneously, or working on several vessels simultaneously at different or overlapping stages of the *chaîne opératoire*.

12.3.1 Raw Materials

The main materials for pottery are water and clay, which together provide plasticity, and one or more fillers (temper). Clay minerals result from the weathering of silicate rocks that contain alumina (Al_2O_3), especially in the form of feldspar (Montana 2017; Rice 2015: 35–40). Among the clays are kaolinite, halloysite, montmorillonite, and illite, but clay sources also include non-clay mineral inclusions. Clay is not necessarily wet and plastic when a potter acquires it; often it just looks like rock and needs to be pulverized and mixed with water to become plastic.

Potters often blend temper into clay to prevent vessels from collapsing under their own weight during manufacture, to control shrinkage during drying, and to alter the characteristics of the fabric, such as porosity, response to rapid temperature change, or tendency for crack propagation. Typical tempers are minerals, such as quartz or dolomitic sand, crushed flint, chert or basalt, crushed mollusc shell, and organic fibers such as wheat chaff, seaweed, or animal hair. Since clay sources frequently contain non-clay inclusions, it is not always obvious which inclusions in a sherd or vessel were intentionally added as temper. A substantial aspect of work on archaeological ceramics involves identifying and sourcing these materials (Braekmans and Degryse 2017; Waksman 2017; Whitbread 2017).

To color some or all of vessel surfaces, potters use pigments, which can include ferric oxide (yellow ochre), haematite (red ochre), manganese, lead carbonate, and carbon, among many other minerals. Technically, "paint" refers to the application of the pigment, but paints typically contain a pigment combined with fine clay and water or an organic oil that will burn off during firing. For glazes (a thin glass layer), the primary ingredient is silica, but lead oxide, potash, soda lime, wood ash, or salt are included as fluxes and colorants (Hodges 1964: 33–34, 42–51; Rice 2015: 159–162).

In addition to the materials for the pottery itself, potters also need to acquire fuel for firing the vessels and will need certain tools for forming and finishing the vessels. The type of fuel, in conjunction with the type of kiln, ventilation and other factors, will dictate the firing temperature and atmospheric conditions during firing. Tools can include mats or wheels for turning vessels, "ribs" for thinning walls (Fig. 14.9), cloths for wetting vessel walls, and various tools for trimming and making incisions and impressions (Hamer 1975).

12.3.2 Processing the Raw Materials

Typically, potters need to pound hard, dry clay to pulverize it and remove contaminants, such as roots and non-clay rocks. Sometimes they sieve the pounded clay with a basket or cloth. The next step is slaking, when the potter adds considerable water to the clay in a shallow pit or vat, disintegrates any remaining lumps, and stirs the pit to put clay particles into suspension. Especially when relatively fine sieves were not available, potters allowed coarser, heavier particles to settle in the water—a process called levigation—while the smaller particles, still in suspension, were decanted off. Some varieties of levigation, with multiple decanting or a sequence of settling vats along a sloped trough, can yield particle sizes less than 50 μ (Rye 1981: 36–37; Rice 2015: 133; Sinopoli 1991: 16). In the final vat, the levigated clay is left until most of the water evaporates and the plastic clay is ready for tempering.

Potters also need to prepare the temper. If it is organic, such as chaff or hair, they chop it to reduce it to small lengths and reduce clumping. Often, the temper consists of rocks, shells, or old pottery sherds that need to be pulverized to an appropriate size, or sand from a beach or stream bed. In both cases, sieving can help select the most useful size range (Gibson and Woods 1997: 27–37; Rye 1981: 37).

12.3.3 Preparing the Body

The potter blends suitable proportions of temper and clay, usually by adding temper to a volume of clay that has been wetted just enough to make it plastic, then kneading on a flat surface or in a shallow bowl. An alternative method sometimes used to prepare large amounts is to add the temper to the clay while it is still suspended in water and then allow the water to evaporate (Rye 1981: 38–39).

The tempered clay, called the "body," then needs thorough kneading or **wedging** to remove air bubbles and ensure that the distribution of moisture and inclusions is uniform (Rice 2015: 133–134). Thorough kneading leaves few voids in the finished pottery except those left by organic temper that burns out.

12.3.4 Constructing the Vessel

The *chaîne opératoire* of pottery typically includes what archaeologists describe as primary and secondary forming, used to create a vessel's rough-out or preform, as well as finishing (Hamer 1975). While these are described separately below, it is not unusual for potters to combine techniques in the same vessel or its parts and some techniques can be involved in both primary and secondary forming. This blurs the distinction.

12.3.4.1 Primary Forming

With materials available and prepared, the first step in building or forming a vessel is called primary forming, which includes several ways to make hand-made pots as well as wheel-throwing and moulding (Rice 2015: 135–145; Rye 1981; Sinopoli 1991: 17–27). A **rough-out** is an unfinished vessel that does not yet have its final geometry, while a **preform** has the intended geometry, but may yet be subjected to finishing. In some cases, primary forming is all that is necessary to create a preform, which may also be the final product.

Pinching involves taking a ball of tempered clay and squeezing it with fingers into a bowl shape, rotating the piece and continuing to squeeze at intervals to thin the vessel walls. The energy source is the potter's hands, applied with discontinuous pressure, to a single mass of clay body. It is typically used to make small hemispherical bowls or the bases of round-bottomed vessels.

Drawing is a similar technique whereby the potter drives a fist into a lump of clay body, squeezes and pulls it upwards from the middle, turns the piece and continues to squeeze and draw the clay body upwards to thin the walls by stretching. This results in a taller and possibly larger bowl than by pinching.

Coiling is a very common primary technique for hand-forming vessels. Its energy source is still the potter's hands, applying pressure discontinuously, but now with assembled parts rather than a single clay body. Typically, the potter begins by patting a ball of clay body into a disk that will serve as a flat base, sometimes setting it on a small mat or shallow basket to facilitate turning the vessel. Alternatively, the potter may begin by pinching a round base and setting it into a bowl or a hollow in the ground to support it during vessel construction. She or he then rolls some clay body into a snake-like cylinder or a taller, thick ribbon, and presses it along the circumference of the base so as to attach it securely. Doing this with one or two coils usually completes the circumference, and the potter then repeats the process with additional coils pressed firmly onto the top of the previous coils, often smearing the outer and inner surfaces of the new coil downward to overlap the lower coils so that they are well attached. By continuing this process with more coils, the potter can increase the height of the vessel wall and also somewhat control vessel shape by inclining new coils inward or outward to alter the diameter (Fig. 12.6). After adding several coils, it is usually necessary to let the vessel rest for a while and partially dry before adding further coils, in order to avoid collapsing the vessel walls, however, the unfinished vessel must not be allowed to dry completely. Alternatively, the potter can make a vessel with a round or pointed base by beginning with a coil that will be the rim of a bowl, and build it upside-down with coils of gradually decreasing circumference that eventually close in. Potters can even build large coil-built pots by using both these approaches to make the lower and upper parts of a vessel separately and then join them together (Fig. 12.7).

Because lower coils need to stiffen somewhat before the potter is able to add further coils, potters may make multiple vessels at once, turning from one vessel to another and then returning to the first vessel to build it up further. One of the implications of this is that builders of coil-made pottery will almost never make just one or two vessels in a potting session.

Slab building involves assembling vessels by joining together flat pieces created by rolling out clay body on a

Fig. 12.6 Views of a traditional potter making coil-built vessels in a village in northern Jordan in 1995 (photos by E. Grossmith): (**a**) forming a disk base, (**b**) the lower parts of two vessels, (**c**) building up the walls of a jar (**d**) bringing the upper part of a jar inward, (**e**) beginning the base for a new vessel while others are drying, and (**f**) joining a pedestal base onto a bowl

surface and cutting it to a desired shape. Like coiling, it employs only the energy from the potter's hands, applied discontinuously. The first slab is used for the base, then other slabs are joined onto the base edge-wise, somewhat like laying mud-bricks in a wall, but pinched together at their edges to form seams that ensure they are well joined. Once the first row of slabs has dried just enough to allow it to support some weight, the potter can add another row, as with coil building, up to the rim of the vessel. Although slab-building can be used to construct all or part of many vessel types, it is particularly suitable for building rectangular and cylindrical vessels, boxes, and very large, immovable containers, such as ovens and grain silos.

Molding in antiquity usually involved pressing a slab of clay body into a mold that had been dusted with a parting agent (fine sand or clay powder). Once removed, the body retained the shape of the mold and could be used as is or joined with other mold-made pieces to make more complicated shapes. Consequently, the energy source in pre-modern molding was still the potter's hands, applied discontinuously, but some mold-made vessels were made from a single mass, while others were assembled from two or more mold-made pieces. The simplest use of molding prehistorically was probably to make bowls by pushing clay body into a small pit in the ground, or into a previously made bowl. This method can also be used to make the base of a round-bottomed vessel that the potter will finish with coiling or slabs.

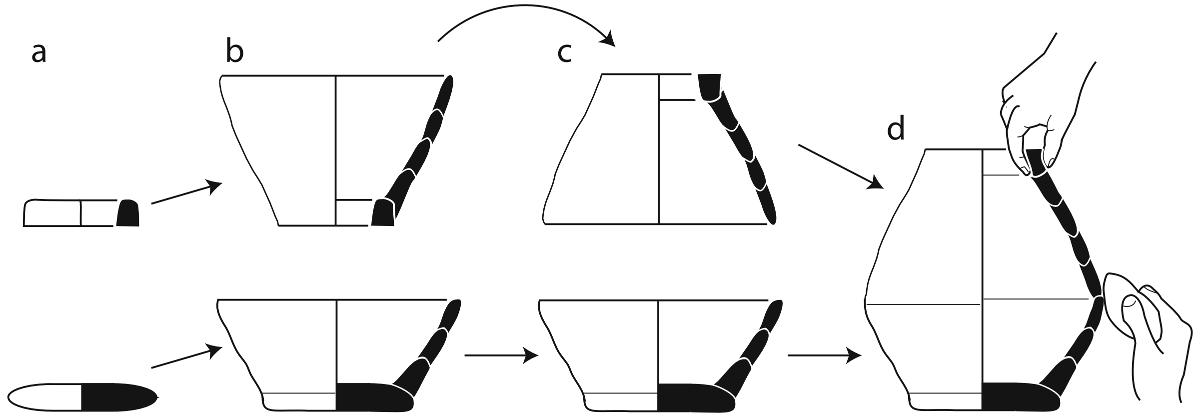


Fig. 12.7 Joining two coil-built sections together to make a biconical or "carinated" jar. The lower part begins, as usual, with a disk base but the upper part with only a ring (**a**), then the vessel walls are built up to make two bowl-like parts (**b**). After flipping over the upper part, the potter can place it on the rim of the lower bowl, carefully join the rim coils, then smooth and trim the surface at the join. The potter can then shape and smooth what is now the rim (**d**) to make a "holemouth" jar, or add a third, smaller section on top for a neck. This method is particularly common for larger coil-built and mold-made vessels

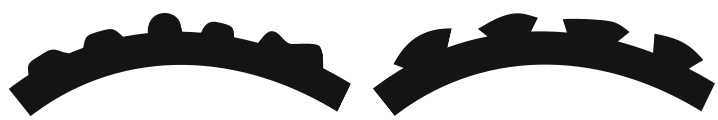


Fig. 12.8 Cross-section through a portion of a complex, mold-made vessel (left) with relief that tapers in such a way as to permit easy removal from the mold. Cross-section at right, by contrast, has features that would prevent extraction from the mold

In modern times, a more common method has been to slosh a clay slurry (clay suspended in water) around in the mold so that clay adheres in a gradually thickening layer on the mold's interior surface, while the porous mold draws water away from the slurry.

One of the important characteristics of molds, especially if they have complicated relief, is that shapes and relief must taper in such a way as to allow removal from the mold (Fig. 12.8). Molding is well-suited for creating shapes that are complex or have relief, as long as the potter keeps the practical constraint of removing the molded pieces in mind.

Wheel-fashioning involves a combination of coiling and rotational kinetic energy. The potter would begin building the vessel by making and joining coils, but would do so on a simple wheel that allowed her to use both hands and the rotation of the wheel to keep the vessel centered and fairly circular in transverse plan, to thin and shape the walls, and to make the coil joins much less visible (Roux and Courty 1998). In this case, the wheel is not simply used to finish the vessel but is integral to shaping a rough-out or preform. As with wheel-throwing, it is necessary to keep the vessel walls moist during shaping and thinning. There are several varieties of wheel-fashioning, differing in whether the wheel is only used to do the final shaping, or is also used to thin the walls, or to help join the coils.

Wheel-throwing is qualitatively different from all the hand-forming techniques and molding, while bearing some similarity to wheel-fashioning in its use of rotational kinetic energy (Rice 2015: 140–145; Courty and Roux 1995). While a single mound of clay body rotates on the wheel, the potter centres it by pressing inwards with wet hands, resulting in a volcano-like cone. The potter then thrusts one hand into its centre and "opens" the cone by drawing material upward and outward to form a bowl shape. The main differences from pinching and drawing is that the potter's hands serve mainly as barriers to the outward thrust of the clay body while it is under the centrifugal force caused by rotation, and the application of the hands is mainly continuous. The main difference from wheel-fashioning is that throwing involves a single mass of clay body and all the forming and thinning, not just some, involves rotational energy. Squeezing and drawing the material upwards, either between thumb on the inside and fingers on the outside, or between two hands, is called "lifting." This thins the walls and increases their height. It is usually necessary to wet the body with water to lubricate it. Moving the hands inward or outward further shapes the vessel by increasing or decreasing its diameter at a particular height. Applying inward pressure with the hands during fast rotation, a process called "collaring," narrows the vessel to form a neck or to close it in completely, as when making a

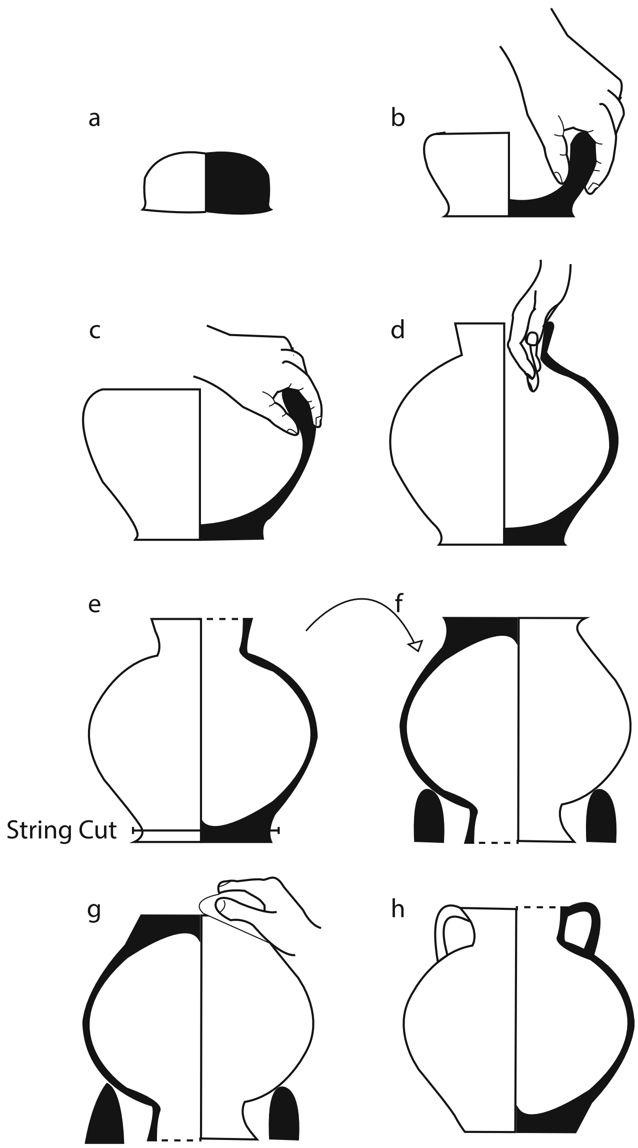


Fig. 12.9 A reconstructed construction sequence for a wheel-thrown cooking pot. (W. Wadsworth, after Franken and Kalsbeek 1975)

base on an upside-down vessel (Figs. 12.9 and 12.10). Shaping involves changing the vessel diameter anywhere on its profile without significantly altering its height. Folding is shaping a rim by flaring it outward or bending it inward and rolling it down to double the rim's thickness or create a complex profile. When the potter has finished shaping the vessel (or portion of a vessel made in parts), she or he cuts it off the hump of clay body by drawing a string or wire just under the intended base while it is still spinning. Turning a new vessel upside-down on the wheel and centering it can allow the potter to form the base, such as a ring-base or concave base, but sometimes potters left the base just as it was cut away.

Wheel-throwing is an early example of mass production, as it is capable of forming many vessels rapidly, but sequentially, unlike the slower but potentially simultaneous production of coil-built pots.

12.3.4.2 Secondary Forming

Secondary forming often follows the initial rough-out of a vessel's shape to refine it or to complete large or complex vessels. If the vessel requires secondary forming or finishing, often the potter needs to let it dry at least a little, to prevent the vessel's collapse, before proceeding to the next step. Several of the techniques described below involve use of formal or informal tools (Van Gijn and Lammers-Keijsers 2010).

Large, complex vessels are often made in several pieces that the potter needs to combine by **joining**. The join results in a thickened seam, similar to the joins in slab-built vessels, although potters typically try to smooth or scrape away any obvious traces of these seams. Many large, globular vessels are made by joining two bowl-like pieces at their "rims," the join often occurring at a "carination," or sharp change in vertical profile (Fig. 12.7). Any of the primary forming methods described above can serve to make the parts to be joined.

Beating, paddling, or hammer-and-anvil, is a technique to thin and even out the walls, especially of coil- or slab-built vessels, after they have partially dried to "leather" hardness. It has the effect of improving bonding between coils or slabs and allowing potters to add texture to the exterior surface. Beating works by placing an "anvil"—a stone or tool, or just the potter's fist—against the interior of the vessel wall while beating the exterior wall repeatedly with a stick or paddle. Using a paddle carved with a pattern or covered with cloth or cord is an effective way to add decoration or texture to the vessel's exterior. Beating is particularly effective for the secondary forming of medium-sized, globular vessels (Rye 1981: 84–85; Rice 2015: 147).

Scraping is a method to thin vessel walls by removing material with a tool held almost perpendicular to the surface, usually when the vessel is somewhat plastic but nearly leather-hard. Sometimes the potter uses a tool with a serrated edge, such as a clam shell (Rice 2015: 147).

Trimming or fettling, somewhat like scraping, involves cutting material away but the tool, often a knife, is held at an acute angle to the surface while the clay body is leather-hard. A common use of trimming is to remove the raised traces of seams, especially on mold-made or other vessels joined from two or more parts (Rice 2015: 147).

Turning is a form of trimming analogous to wood-turning on a lathe, and the only major secondary forming technique that involves rotational kinetic energy. Instead of trimming the vessel freehand, the potter holds a tool at an acute angle to the surface while rotating the vessel on a wheel. This removes long curls of material from the vessel walls in a nearly continuous spiral and is very effective at thinning the walls and disguising or removing traces of earlier forming processes.

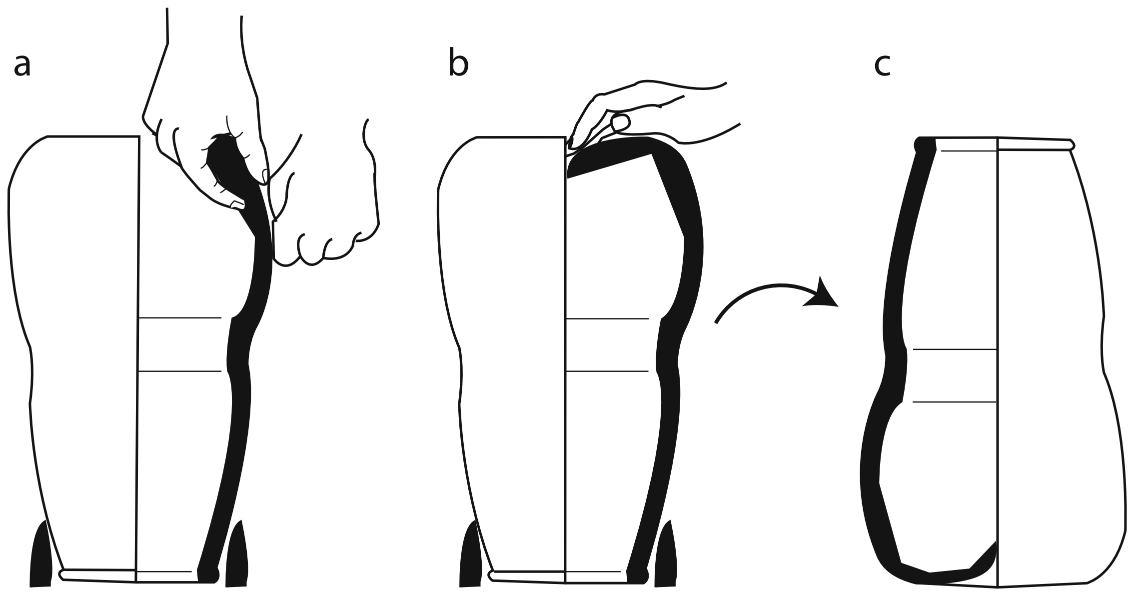


Fig. 12.10 Closing the base of a tall jar by turning it while upside-down: "closing" the vessel by drawing the upper part of the wall inward (**a**), depressing it to make a concave base (**b**), and the finished, upright vessel (**c**). (W. Wadsworth, after Franken and Kalsbeek 1975)

12.3.5 Finishing Techniques

Once formed, preforms of vessels can undergo one or more techniques to alter their surface characteristics. These can involve displacing or impressing material at the surface, applying thin coats of slip, paint, or other sealants, or removing small amounts of material from the vessel by cutting or carving.

Smoothing and **burnishing** slightly displace material near the vessel surface. Both involve rubbing the surface while the vessel is leather-hard or dry, but not yet fired. Smoothing with hands or a tool is less aggressive, can be done with or without rotation on a wheel, and leaves a matte surface. Burnishing, often performed after the vessel is slipped, involves rubbing the surface with a smooth pebble to compact and reorient the clay particles near the surface. This results in a hard, reflective and less porous surface (Rice 2015: 148). "Pattern burnish" involves selective burnishing to make decorative patterns on the surface, such as net patterns.

Impressing involves greater displacement of the material in the vessel by pushing a tool with considerable pressure against the surface while the clay body is still somewhat plastic or nearly leather-hard. The tool can just be fingertips or fingernails, the edge of a shell (e.g., David et al. 2012), or a stamp or seal (stamping). "Punctate" decoration is a type of impressing with the ends of sticks, bones, canes, thorns, porcupine quills or other narrow objects.

Rouletting is very similar to impressing except that the tool is cylindrical and is rolled over the vessel surface. The tool can be a carved stick, a stone cylinder seal, a gear-like wheel, or a cord-wrapped stick (Rye 1981: 92–93; Hurley 1979). Rouletted patterns repeat over patches of the vessel surface or even over its whole circumference. As with beating with a cord-wrapped paddle, rouletting can result in a "corrugated" or roughened surface that provides good grip and a greater surface area that may even improve heat transfer in cooking pots (Rice 2015: 138).

Combing and **incising** are techniques that somewhat aggressively displace, but do not remove, material from the vessel surface. Incision involves cutting into the surface of a leather-hard or only slightly plastic vessel with a narrow tool, slightly displacing material outwards from the cut, then dragging it along so that it displaces some material toward the final end of a linear or curvilinear groove. Combing employs a tool with multiple prongs, such as a fork, dragged along a slightly plastic surface to displace material in such a way as to create several parallel grooves. Holding the tool still against the surface of a rotating vessel allows a potter to create single (incised) or multiple (combed) horizontal lines on a vessel, while moving it up and down results in a wavy pattern (Fig. 12.17).

Finishing techniques that remove material from the vessel include **piercing** or perforating vessel walls or handles while leather-hard, drilling them after they are fired, and carving areas to shape them or add decoration (Rye 1981: 91–92).

Other finishing methods involve adding material to the vessel surface. While there can sometimes be unintentional surface coatings (e.g., "self-slip" resulting from the potter's wet hands during forming), potters often apply intentional coatings.

Slip is a suspension of clay particles in water. Use of this term to describe a coating on pottery refers to a thin layer of fine clay that adheres to the surface of a vessel and was fired along with it. It may or may not have the same colour as the underlying fabric. Potters sometimes apply slip to only a portion of the vessel surface as a decorative effect. On low-fired vessels, burnishing the slip before firing improves the slip's adhesion and gives the vessel a lustrous surface (Rice 2015: 162–164).

Paint is a material that the potter applies to the vessel surface before or after firing. Some authors (e.g., Rye 1981: 40) argue that paint can be a slip or pigment, defining paint by the potter's intent rather than the materials. However, most authors restrict the term "paint" to refer to the application of pigments, media containing colouring agents such as metallic oxides and sometimes oils, clay or other organic materials that make it easier to brush the colour on or make it adhere (Shepard 1961: 177; Rice 2015: 161–162). Some pigments are mixed with significant amounts of clay to improve their flow and ability to adhere, making the distinction between slips and paints somewhat arbitrary. Potters usually apply paint to dry surfaces so that the vessel walls will quickly absorb moisture in the pigment, resulting in crisper edges to the painted design (Rye 1981: 41).

Unlike paint, a **glaze** is a thin coating of glass fused to the vessel surface during firing. Most ceramic glazes consist of silica (SiO_2) and fluxes, such as potassium, sodium, calcium, magnesium, and lead, that bring the melting temperature down from more than 1700 °C to 900–1300 °C (Rye 1981: 44–46). Glazes often themselves add color, but can also be applied over paint or slip.

A **resist** is any material that the potter applies to the surface to prevent the adhesion of slip, paint, or glaze, so that some regions of the vessel surface are left without those coatings, as a decorative device. A wax resist burns off during firing, while resists made from leaves, paper or other materials can be stuck onto the vessel and then removed after the slip or paint has been applied but before firing (Shepard 1961: 206–213; Rye 1981: 43–44).

Other coatings that someone applies after a vessel's firing, often as a **sealant**, include lime or pitch, while vessels can also acquire surface residues during use.

Appliqué is a finishing technique that involves joining shaped pieces of plastic clay body to a nearly leather-hard surface by pressure. The applied piece can be functional, such as a handle, decorative, such as a modelled figure, or both, such as "rope" decoration that has an aesthetic aspect but also makes the vessel easier to grasp or move around and obscures seams (Rye 1981: 93–95).

12.3.6 Drying and Firing

Carefully drying the formed vessels is essential to avoid their destruction in the kiln by somewhat explosive expulsion of steam. Drying must be gradual and can take days or weeks, depending on weather conditions and the type and texture of inclusions. Using artificial heat sources to warm the vessels can help them to dry under humid conditions, while some potters will increase the firing temperature very gradually to ensure the escape of moisture from vessel walls before subjecting them to high temperatures (Gibson and Woods 1997: 46–47; Rice 2015: 171–177).

Firing the vessels at temperatures above about 500 °C breaks down clay minerals and transforms them into a ceramic material. The duration and atmosphere of firing, not only the temperature, affect the physical characteristics of the resulting ceramic (Rice 2015: 99–116). These range from relatively poorly fired earthenwares, with mainly crystalline and porous fabric, to porcelains, with predominantly vitreous, translucent, and non-porous fabric (Table 12.2). Open firing is sufficient for production of earthenwares, while kiln-firing is necessary for stonewares and porcelains.

In **open firing**, dried vessels are arranged on top of fuel, usually in a shallow pit (sometimes loosely described as a "kiln"), and more fuel is intermingled among them and piled around them. Spaces between fuel permit better airflow to attain temperatures above 800 °C, but both temperature and atmosphere can vary considerably. The fire reaches maximum temperature within about 20 minutes and rarely exceeds 1000 °C even in the hottest places. Temperature declines as the fire dies down and then rises again when someone adds or rearranges fuel or the draft changes. Typically, vessels with wide apertures are fired upside-down; if ash accumulates, this insulates the rims, allowing them to cool more slowly to avoid cracking (Rye 1981: 98), and may cause the interior and rim to be reduced (grey or black) while most of the exterior surface is oxidized (red or yellow). It is difficult for

Table 12.2 Characterization of ceramic bodies by firing temperature

Body	Porosity (%)	Firing Temperature (°C)	Characteristics
Terra-cotta	> 30	600–1000	Coarse, porous
Earthenware	10–25	900–1200	Non-vitrified
Stoneware	0.5–2	1200–1350	Vitrified
China	< 1	1100–1200	Vitrified
Porcelain	<< 1	1300–1450	Hard, white, fine, translucent

Modified from Rice (2015: 5)

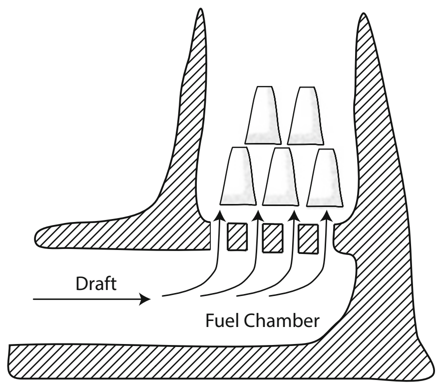


Fig. 12.11 Cross-section through an updraft kiln (Y. Salama)

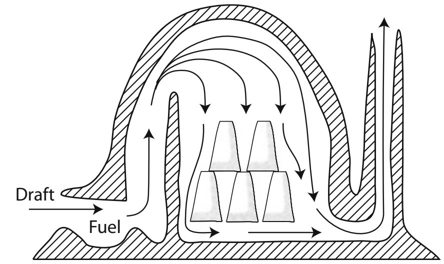


Fig. 12.12 Cross-section through a downdraft kiln (Y. Salama)

open firing to produce a reducing atmosphere except in these small areas, however, although spots of dark "smudging" can occur in other places where access to oxygen was restricted. Open firing owes its success to the fact that the most pronounced physical changes in vessel walls, including loss of water, organics, carbonates, sulphates and sulfides, occur at temperatures well below 900 °C and even sintering, vitrification, and collapse of the clay's crystal structure occur in the range of 900 to 1000 °C (Rice 2015: 100–115).

Kiln-firing separates fuel from the pottery. Combustion of fuel occurs in a firebox and the hot gasses pass through flues into a firing chamber that contains the vessels. Kilns allow potters to achieve temperatures as high as 1300 °C, to control the rate of heating, and to manipulate the atmosphere by disrupting or increasing the flow of oxygen.

An **updraft kiln** (Fig. 12.11) has the firing chamber above the firebox so that it resembles a chimney. The potter controls airflow at the stokehole. Mayes (1962) notes that it is difficult to create a reducing atmosphere in updraft kilns, although potters can darken the pottery by blocking the stokehole and letting the fuel smoulder during cooling (Rye and Evans 1976). Updraft kilns distribute heat rather unevenly (Rye 1981: 100).

Downdraft kilns achieve temperatures around 1300 °C, and provide good control over atmosphere, so they are suitable for making porcelain and reduced stoneware. Hot gasses from the firebox flow upward over a "bag-wall" and then downwards through the chamber and up again through flues and a chimney (Fig. 12.12). The chimney is necessary to draw the flame over such a long distance (Rye 1981: 100).

In any of these firings, potters could determine whether vessels have reached the right temperature by their color or by using "draw trials," small pieces with a purpose similar to modern pyrometric cones (small cones made of materials similar to glazes that signal "work heat" by their deformation; Rice 2015: 111) that the potter can remove to see how well they have been fired and whether they are reduced or oxidized.

As noted in connection with open firing, all these types of firing cause the material to undergo chemical changes that radically alter its character, and then the ceramic fabric cools to result in a new material that conventionally falls into one of several categories (Table 12.2). In both low-power examinations and petrography, observers can identify inclusions that the potter may have added to the paste (Kempe and Harvey 1983; Nesse 1986), or determine what phase changes minerals have experienced, while the methods from analytical chemistry can sometimes help identify clay sources and provide a different way to assess some of the inclusions, since the chemistry of minerals is well known. As a bonus, the former methods as well as micro-CT scans and xeradiography can help us assess the density and orientation of inclusions (Applebaum and Applebaum 2005; Jizba 1971).

12.3.7 Use, Re-use and Recycling

We should not overlook the last stages of the *chaîne opératoire*. Considerable archaeological research focusses on identifying the uses of pots, while even broken pots were not only discarded, as people sometimes exploited the sherds of broken vessels to make tools. They could drill holes in their center and chip their edges to make weights, spindle whorls, and gaming pieces, use unmodified or slightly modified sherds as scrapers or "ribs" for pottery making (Vieugué 2015), or grind up sherds to use as temper in subsequent pottery-making; this type of temper is called "grog."

12.4 Pottery Attributes

12.4.1 Attributes for Identifying Raw Materials

Many of the attributes that archaeologists examine in pottery pertain to choices of raw materials and their preparation, while others are clues to forming and finishing techniques or firing temperatures and atmospheres. Some of the key methods for this include examination of fresh breaks (transverse or radial sections) under low-power magnification, petrography using polarizing-light microscopy, and chemical characterization by X-ray fluorescence, neutron activation or other analytical methods.

While various chemical analyses, such as Instrumental Neutron Activation Analysis (INAA) and portable X-Ray Fluorescence Analysis (pXRF) can be helpful in identifying clay sources (e.g., Lyons et al. 2018), archaeologists make substantial use of optical petrography to identify inclusions, which may pertain to tempering "recipes," in the fabric of sherds.

Optical petrographic techniques exploit the ways that light interacts with crystalline and non-crystalline materials in thin sections—slices through the pottery mounted on glass slides and polished down to 30 μm (microns or micrometers)—to identify the minerals their fabrics (Nesse 1986; Quinn 2013).

The use of polarized light, by passing normal light through a polarizing filter, allows us to distinguish various minerals in thin section more easily. When polarized light passes through "anisotropic" minerals, it is split into two rays, called "fast" and "slow" rays, and "birefringence" is the difference between these rays' angles of refraction. These two rays can also interfere with one another, leading to notable changes in the color of minerals in polarized light. But, when the vectors of these rays are at right angles to the angle of polarization, the mineral they are passing through appears dark, and the angle between the mineral's cleavage angle and the dark position, found by rotating the microscope stage, is called the "extinction angle." Some minerals, such as garnet, are isotropic, which means that they look the same no matter what the angle of the polarized light. Extinction angles and the presence of isotropy thus help us identify minerals in pottery fabrics.

12.4.2 Attributes for Identifying Forming and Finishing Techniques

Owen Rye (1981) provides a good summary of attributes, mainly observable with the naked eye or low magnification, that can help us identify the primary and secondary forming techniques and finishing techniques that were involved in making vessels (see also Courty and Roux 1995; Rice 2015; Roux and Courty 1998). Some more recent methods, including microscopic and chemical analysis, can sometimes supplement these. Note that subsequent secondary forming or finishing can obscure or eliminate some of the traces noted below.

Pinching can leave traces as regularly spaced, shallow indentations left by finger or thumb, especially on the interior. These tend to have a vertical orientation and a length not exceeding that of a thumb or fingertip. In horizontal (transverse) section, we see an almost sinusoidal waviness in surface profile unless secondary forming later removed it. Inclusions tend to be drawn upward by squeezing, resulting in slight preference for vertical orientation in a radial section, a very slightly vertical but mostly random orientation of their long axis in tangential view, but a tendency to be parallel to the vessel surface (Fig. 12.13; Rye 1981: 52).

Drawing similarly results in a series of finger-width indentations around the pot's circumference, oriented vertically and generally longer than in pinching. Again, later finishing may obscure or remove these features, but may

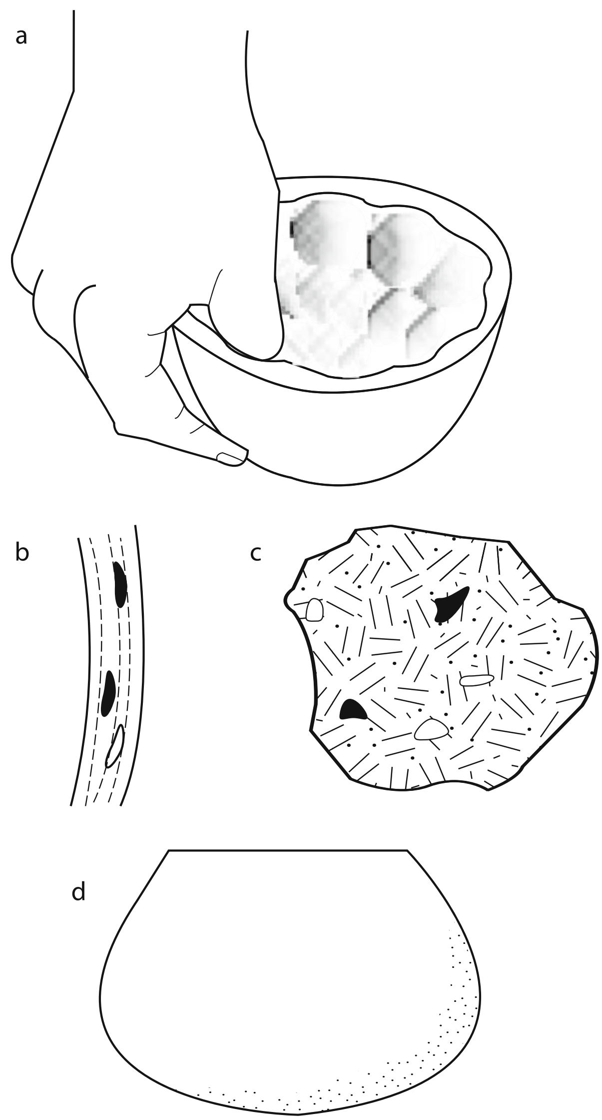


Fig. 12.13 Physical traces of pinching (after Rye 1981). Fingers have left concave impressions on interior (**a**) and exterior with somewhat vertical orientation, and elongated inclusions have a slight preference for vertical orientation in radial section (**b**) and tangential view (**c**). Pinching often results in small, hemispherical or ovoid vessels (**d**)

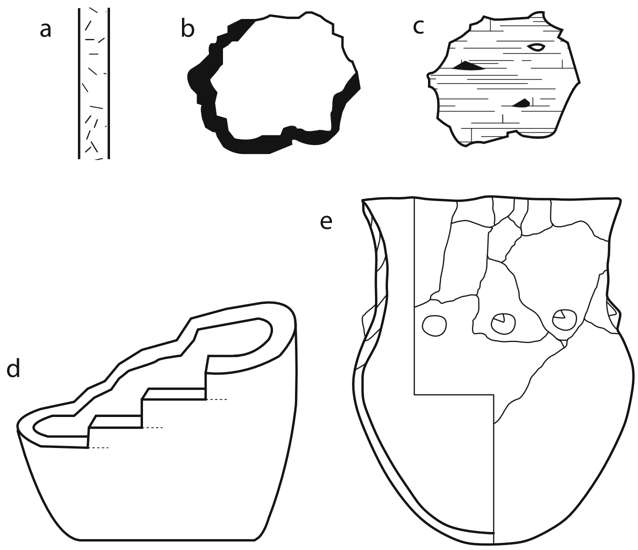


Fig. 12.14 Evidence of coil-building (after Rye 1981) includes variable or undulating wall thickness, tendency toward irregular or step-like fracture along coil bonds (**d**), and preferred horizontal orientation of elongated inclusions in tangential view (**c**), but random orientation in radial view (**a**). In some cases boundaries between coils may be visible in radial section (**e**) (W. Wadsworth and Y. Salama, after Rye 1981, and Pétrequin and Pétrequin 2017: 21)

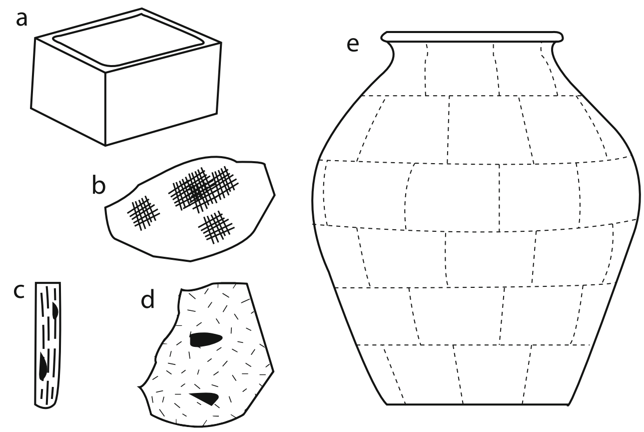


Fig. 12.15 Evidence for slab-building (after Rye 1981) can include non-circular shape (**a**), traces of scraping or combing at seams (**b**), tendency for elongated inclusions to have random orientation in tangential view (**d**) but to be parallel to interior and exterior surfaces in radial and transverse views (**c**), and presence of vertical and horizontal joins (**e**) that may be particularly evident in X-rays or CT scans

still leave somewhat wavy variations in wall thickness in the transverse plane. In radial section and tangential view, inclusions more noticeably prefer a vertical orientation of their long axis than in pinching.

Coiling, unless finishing obscures its traces, results in periodic variations in wall thickness in the radial plane. Sherds may also tend to break along coil boundaries, sometimes in meandering or step-like fashion, and sometimes leaving a rim-like edge or "false rim" (Fig. 12.14; Callander 1937; Gibson and Woods 1997: 39–42; Rice 2015: 135–137). Mineral inclusions can be relatively large and even angular. Elongated inclusions and voids have a tendency toward horizontal orientation in tangential view, while radial sections tend to show inclusions end-on and with apparently random orientation (Fig. 12.14; Gibson and Woods 1997: 37–42; Rye 1981: 67–69).

Slab-building's slabs can exhibit very smooth surfaces as a result of being rolled out but there are usually many surface markings from scraping or combing the seams in an attempt to obliterate them, increase adhesion, and even out thickness. If the slabs are rolled under sufficient pressure, this flattens inclusions so that they are parallel to the surface in radial and transverse sections but randomly oriented in tangential view. Fracture sometimes occurs along seams, but rolling can also create a laminar structure that allows flaking off of parts of the surface after fracture in a plane parallel to the surface (Fig. 12.15; Rye 1981: 71–72).

The complex relief of some **mold-made** ceramics is the first and sometimes sufficient indication of their primary manufacturing technique, but mold-made ceramics often also exhibit traces of seams from joining parts or the secondary techniques used to remove or disguise seams (Figs. 12.7 and 12.16). Elongated inclusions can sometimes have a slight tendency to align parallel to the surface of the mold, but are generally fairly random in orientation (Rye 1981: 81–83). Potters making molded vessels with complex relief surfaces may prefer inclusions to be small, to avoid imperfections.

Wheel-fashioning yields somewhat different traces depending on how much rotary force contributed to shaping, thinning and building the vessels. Vessel walls can show irregular micro-relief and parallel "blisters" or raised areas, but stretching of the clay body and rilling usually only occur if rotational kinetic energy was used through most of the stages of vessel construction and thinning. Orientation of elongated inclusions tends to be random in radial section and horizontal in tangential view, as with coiling. Joins between coils are less visible than in fully coil-made vessels (Roux and Courty 1998).

Wheel-throwing is generally easy to recognize even though vessels turned during secondary forming or finishing or constructed by wheel-fashioning can sometimes have traces similar (though not identical) to those from wheel-throwing. Opening tends to leave spiralling thumb indentations on the inside of the base and, if the potter does not expect these to be visible on the finished product (generally for closed vessels), they will remain (Fig. 12.17). Lifting not only leaves spiral finger grooves, or "rilling," up the interior and exterior walls of the vessel, but also gives elongated inclusions a tendency toward diagonal orientation in the tangential view, something that does not occur in wheel-

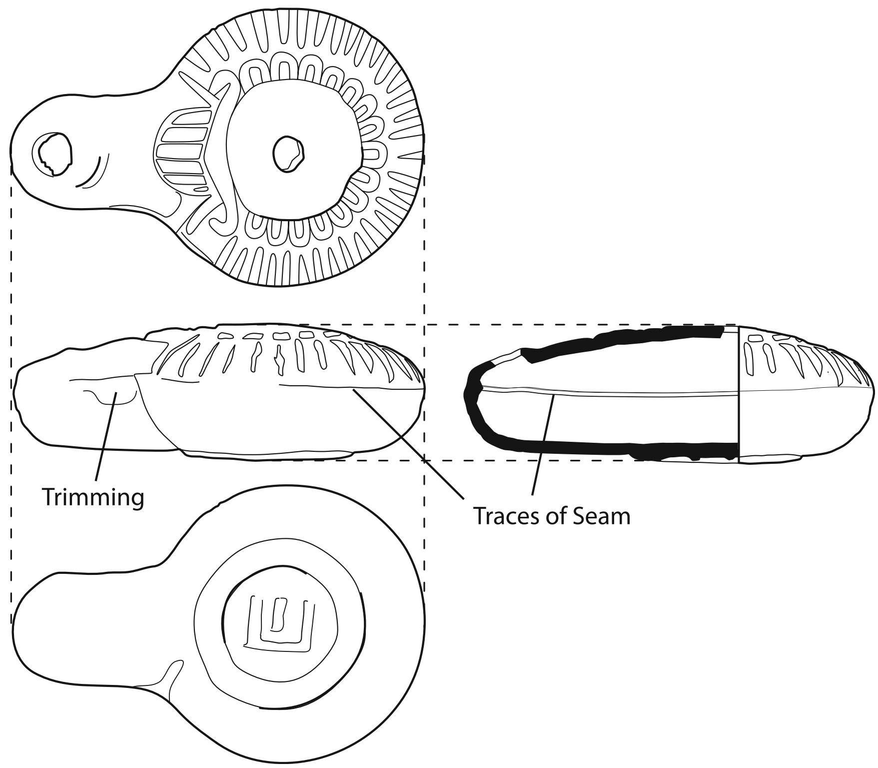


Fig. 12.16 Mold-made vessels, like the Greek lamp shown here, tend to have complex shapes with relief decoration that tapers to permit easy removal from the mold, evidence for trimming excess material from the seam between parts, but excess preserved along interior seams that would not be visible in whole vessels (Y. Salama)

fashioned pots. Removal of a vessel from the hub of clay body on the wheel can result in a "string-cut base that shows spiral or shell-like drag marks from the string (Fig. 12.18) and fingerprints on the lower exterior walls where the potter picked it up, unless later forming or finishing removed these (Courty and Roux 1995; Rice 2015: 140–145; Rye 1981: 75). Often, the lower parts of walls are thicker than upper ones, while there are no very noticeable variations in thickness around the circumference. Wetting the vessel during throwing can leave a "self-slip" that is usually thinner and less regular than an intentional slip. Inclusions in wheel-thrown pots are not strongly angular and typically much less than 1 mm in length when walls are less than 5 mm thick, while thorough kneading and throwing ensure that there are very few and only small voids. S-shaped cracks on the base and spiral fracture of vessel walls are common features. Wheel-thrown vessels are typically very slightly non-circular in transverse view. Finishing of wheel-thrown vessels is also often done on the wheel, so that painted or incised lines are neatly horizontal and regular, sinusoidal, or spiralling (Rye 1981: 75–80).

Beating leaves rounded indentations from the anvil on the vessel interior and smooth or textured facets on the exterior surface from the paddle (Fig. 12.19). This results in regular variations in wall thickness. Use of a wet paddle can also create localized "self-slip." Because beating puts local stress on the leather-hard fabric, there may be star-shaped cracks around larger inclusions at the surface, while cross-sections of vessel walls will show preferred orientation of elongated inclusions parallel to the surface. There can be laminar structure that causes lens-shaped flakes to spall off the surface. Beating is more common on globular, spherical or hemispherical vessels.

Scraping with a smooth-edged tool produces drag marks as larger inclusions are drawn along the surface (Fig. 12.20). Scraping tools can also leave facets on the surface that end abruptly (Rye 1981: 86).

Trimming can also leave drag marks, especially in gritty fabric, but usually results in smooth facets with sharp or torn edges where a curl of material was pulled away. Facets may be elongated and irregular (Rye 1981: 87) but, as with most such techniques, later finishing can obscure these traces.

Turning has traces that are similar to those of trimming, but the drag marks are oriented along the rotation or spiral, and eccentricities in wall thickness are common because the leather-hard vessel was not perfectly centered when it was returned to the wheel.

Smoothing on a "slow" wheel can create horizontal or spiralling traces and, if it is done with wet hands or cloth, can create a "self-slip."

Burnishing leaves narrow, linear or slightly curved streaks of reflective polish on the ceramic surface. Sometimes, burnish strokes alternate in direction to create net-like patterns, while very dense burnish gives the surface an almost metallic appearance.

Impressing or **stamping** is usually quite obvious because the impression retains, in negative, much of the shape of the

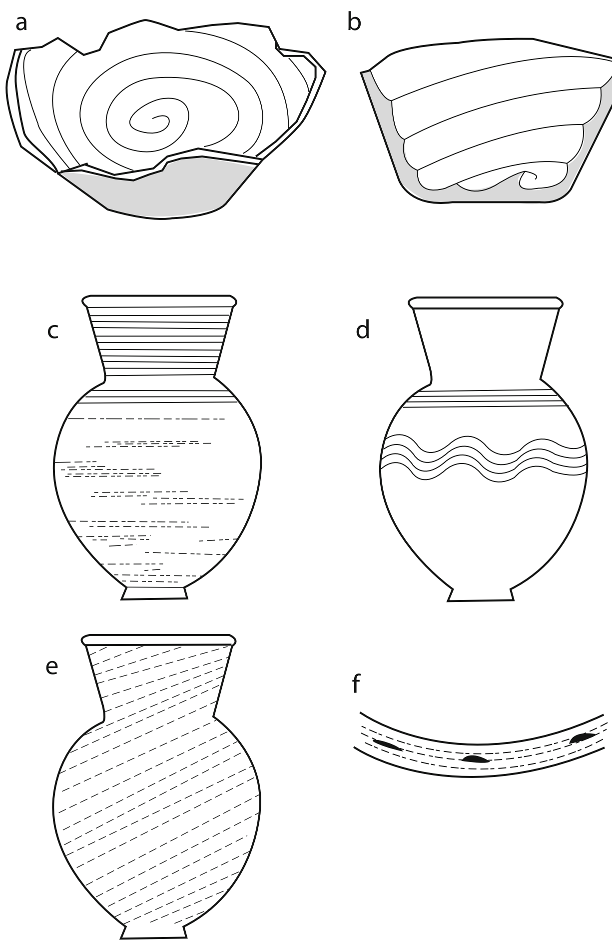


Fig. 12.17 Physical traces of wheel-throwing include spiralling "rills," especially on interior surfaces of bases and jars (**a**, **b**), near-circular plan in transverse view, tendency for elongated inclusions to be oriented parallel to vessel walls in transverse (**f**), but diagonally in tangential view (**e**), while finishing on a wheel may show in incised or painted lines that are horizontal and straight or wavy (**d**). (W. Wadsworth, after Rye 1981)

tool or fingernail used to make it and without the drag marks associated with incision. **Rouletting** leaves very similar traces except that the repetition of the impressed design demonstrates that it came from a rolled tool. Careful examination and experiment can sometimes allow identification of tools, or knot patterns on them, that were used for beating, stamping or rouletting (Figs. 12.21 and 12.22; Hurley 1979).

Because **combing** and **incision** displace material so much, they tend to leave small mounds of material near the end of each stroke and to some extent along its margins. Under magnification, there can also be drag marks along the incision walls.

Slip appears as a distinct layer in cross-sections of vessel walls, while "self-slips" that result from wetting or smoothing the vessel during construction and finishing may not be as distinct. Slips have smooth surfaces, and sometimes show a fine network of hexagonally oriented fractures that do not continue into the fabric (Rye 1981: 54). **Paint** can be difficult to distinguish from slip except by chemical analysis or sometimes color, especially as clay is often a major component. The glassy appearance of **glazes** makes them easy to identify. Archaeologists commonly use Munsell charts to classify the colors of slips and paints (see p. 302), while increasingly using hand-held digital meters for the same purpose.

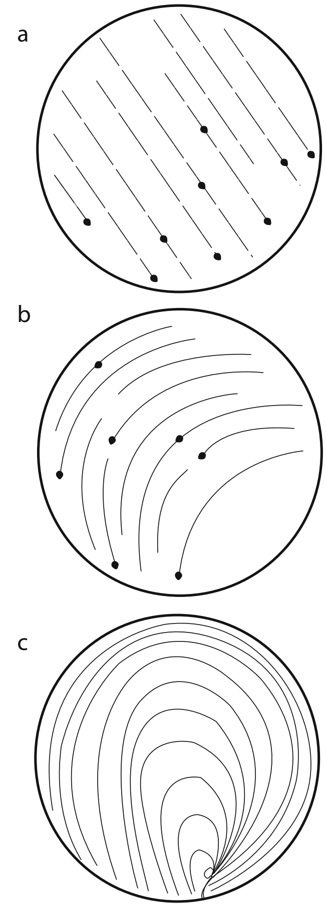


Fig. 12.18 Drag marks on string-cut bases of thrown vessels. (W. Wadsworth, after Rye 1981)

Piercing before the vessel is completely dried displaces material in the direction of piercing, leading to a raised rim around the hole on the opposite side from where the piercing tool was pushed. **Appliqué** is fairly obvious, and often results in insecure joins so that the vessel tends to break along the join. In low-fired earthenwares, air pockets sometimes occur at the boundary between vessel wall and adjoined handle or decorative element.

12.4.3 Attributes for Identifying Firing Conditions

Because firing alters the chemical structure of clay and other minerals, it leaves physical clues to firing conditions, including atmosphere and maximum firing temperature, although not all of them are visible to the unaided eye.

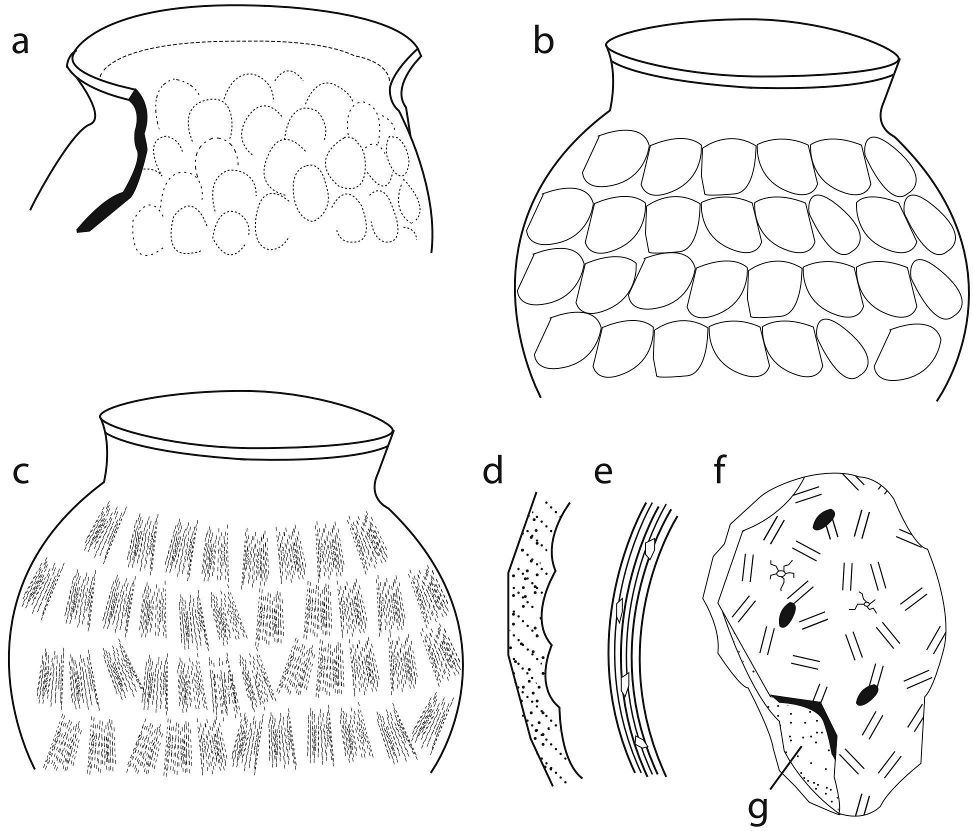


Fig. 12.19 Physical traces of beating (W. Wadsworth, after Rye 1981) include rounded indentations from the anvil on the interior (**a**) and near-flat facets on the exterior surfaces (**b**), and the exterior facets can be textured (**c**), if the paddle was carved or covered with twine or fabric. The facets and indentations also appear in radial and tranverse section (**d**). There can also be small, star-shaped cracks around larger inclusions (**f**), elongated inclusions have preferred orientation parallel to the wall surface in radial and transverse section (**e**), and there may be laminations in vessel walls that tend to flake off

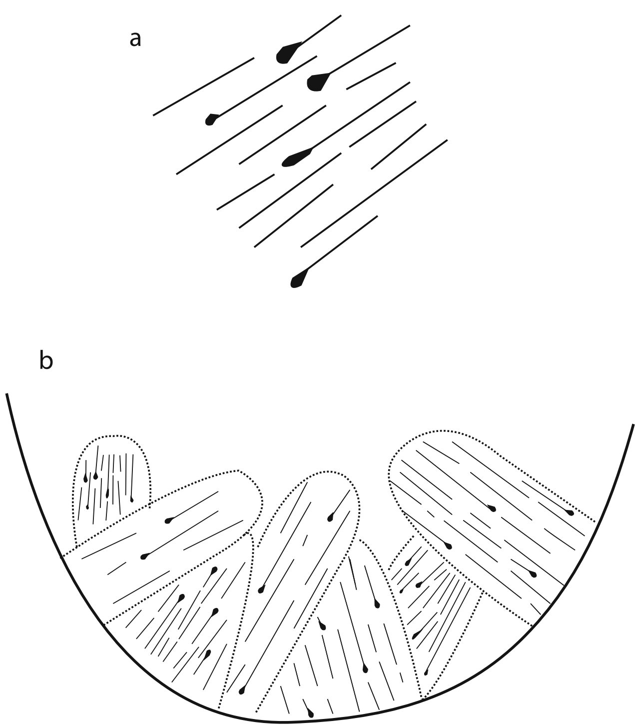


Fig. 12.20 Physical traces of scraping include (**a**) dragging of inclusions in one direction and (**b**) flat, serrated or slightly concave facets. (W. Wadsworth, after Rye 1981)

Determining the maximum firing temperature is called archaeothermometry (Rice 2015: 376–387). Simple visual clues to this temperature can be found in cross-sections through sherds that were fired below 1000 °C. Long firing in an oxidizing atmosphere generally results in a well-fired, light-colored (pink, red or yellow) fabric throughout. However, firing atmosphere, and not only temperature, affects this color. A black or grey core results if oxidation was incomplete, if organics were used as temper, or if there was a reducing atmosphere during firing (Fig. 12.23; Rye 1981: 112–116). However, these clues do not allow precise estimates of firing temperature.

A more dependable use of color for archaeothermometry is to do refirings. After recording the original color of a sherd, it is refired at gradually increasing temperature increments. If, for example, there is no change to the sherd's color when refired to 700 °C, but the color does change at temperatures above 800 °C, we can infer that the original firing temperature was in the range of 700–800 °C (Rice 2015: 378).

Other methods that have been used for estimating maximum firing temperature include identifying temperature-dependent phase changes in the minerals in the pottery (Maggetti 1982), gradually reheating sherds in a lab oven while measuring their thermal expansion and shrinkage with a dilatometer (Kaiser and Lucius 1989; Rice 2015: 380–381;

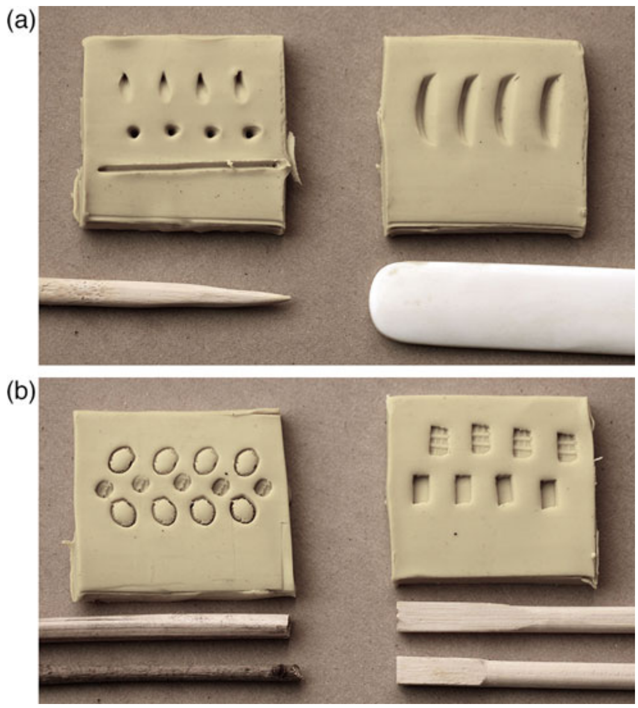


Fig. 12.21 Angled and vertical punctate and incision with pointed stick (**a** left), impression with the spatulate end of a bone weaver's sward (**a** right), impression with reed or blunt stick (**b** left), and punctate with incised or plain, squared stick (**b** right, see also Sampson 1988: 55)

Fig. 12.22 Two examples of rouletting with a cord wrapped around a stick. (see also Hurley 1979)

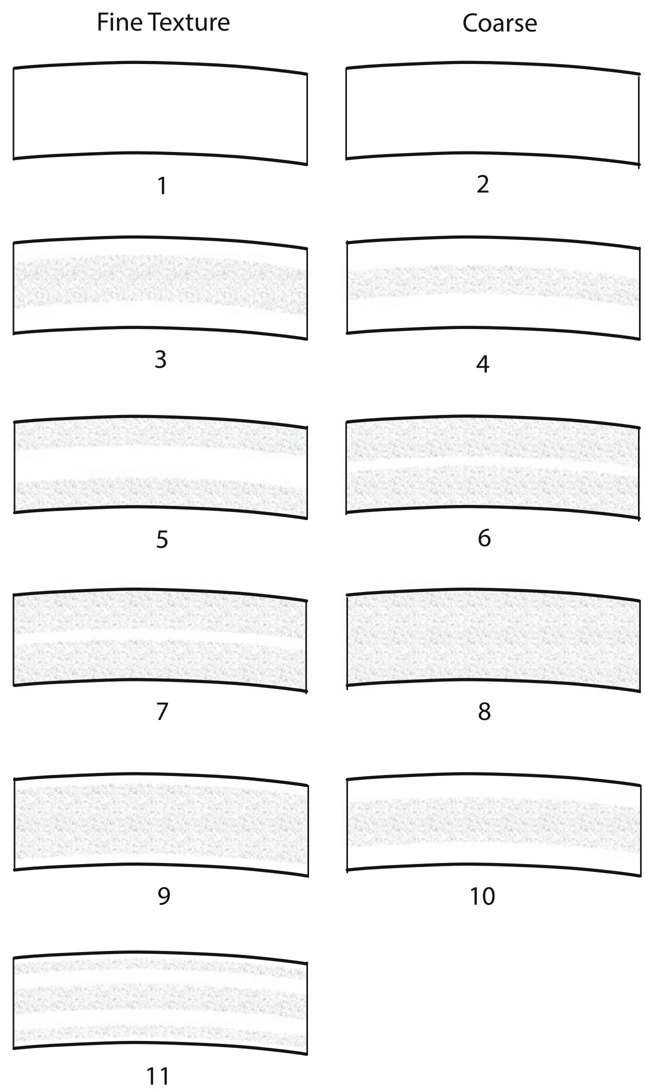


Fig. 12.23 Schematic cross-sections through sherds with fine-textured and coarse clays, and oxidized without organics (1, 2), oxidized with organics (3, 4), reduced (5, 6), reduced with possible organics (7, 8), reduced, then cooled in air (9, 10), and reduced, oxidized, then reduced again (11). (Y. Salama, after Rye 1981: 116)

Roberts 1963; Tite 1969; Zhu et al. 2013), thermoluminescent spectra from the glow that results from energy trapped in quartz inclusions during reheating (Koul et al. 2000), or tracking changes in magnetic susceptibility or paramagnetic resonance of ceramics as they are reheated (Mangueira et al. 2011; Rasmussen et al. 2012; Rice 2015: 383–384).

12.5 Functional Analysis of Pottery

As with lithics, archaeologists are often interested in determining the function of pottery vessels (Rice 2015: 411–432). At its simplest, the primary utilitarian function of most pottery is containment. Like such other containers as pits, silos, glass bottles, boxes and baskets, pots prevent the accidental dispersal of their contents, while still allowing some access for the intentional addition or removal of contents. Even smoking pipes are containers in that the pipe bowl contains and allows introduction of tobacco while allowing removal of ash and escape or inhalation of smoke. Drain pipes also allow introduction and egress of liquid while containing that liquid during the course of its movement from one place to another through a drainage system.

However, archaeologists often want more specific information on function, such as what they contained or were intended to contain. Just as other tools, they could have multiple functions or have been used for purposes that their makers did not intend or perhaps even imagine. A vessel made to contain beer might also be used to serve beer, show hospitality, or display ethnicity, status, or marketing signals, or even to contain loose change or car keys. Some archaeologists distinguish between utilitarian functions, such as mundane use of a pot to contain water, and symbolic or social functions, but any artifact can have multiple functions at once or sequentially.

Most archaeological investigations of function either consider the designed properties of the artifact or traces of its last or habitual use. The former approach is based on the assumption that the designers or makers of artifacts selected design attributes that they thought would make the artifact useful or effective for a particular intended purpose or task, under the constraints that cost and competing design attributes impose (Braun 1983; Henrickson and McDonald 1983; Scholfield 1948; Smith 1985). Sometimes archaeologists use experimental research to evaluate hypotheses based on design principles (Bronitsky and Hamer 1986; Jeffra 2015). The latter involves examining damage and other traces on the artifact, or chemical residues on or in the artifact that could have been left by vessel contents during use.

The design approach anticipates features that might make a pot more useful for a task and then examining pots or sherds for the expected constellation of attributes and contextual associations (Skibo 2013: 27–55). For example, a well-designed cooking pot needs to withstand the thermal shock of sudden heating and cooling, which affects the choice of mineral tempers (Bronitsky and Hamer 1986). It also needs to be large enough to accommodate a typical meal of the type for which it was intended, and have an opening large enough to allow stirring, addition of ingredients, and removal of cooked food. Accommodation for a lid, such as rims that are bevelled or ledged on the inside, may be a good design feature to reduce the escape of heat from the orifice, or retain or control the escape of steam. A round bottom may provide more efficient heat transfer from an open flame if the pot is suspended or perched on stones or a tripod, or a flat bottom may be optimal for heat transfer from a stove top, and cooking pots have a tendency toward generally globular shape (Fig. 12.24). Handles are useful for manipulating the pot without burning one's hands while it is hot.

The design features of a water jar would be quite different. They might include a low center of gravity and perhaps a flat base (if intended for placement on a table or hard floor) to avoid spillage, a restricted orifice to decrease evaporation from the water surface, but, in hot climates, a fabric that is porous to encourage evapotranspiration from vessel walls to keep the water cool, or slipped surfaces to make walls more impermeable to prevent water loss. For large water jars that are rarely moved, there might be no handles; otherwise, handles are likely to be high on the vessel to provide leverage while moving them.

Vessels for transporting liquids and other materials would have still other characteristics. A literally classic example is the Greek amphora. Amphoras were used to transport wine, oil and fish pastes by sea, and they are usually tall, with a high center of gravity, narrow neck flanked by two handles, and a

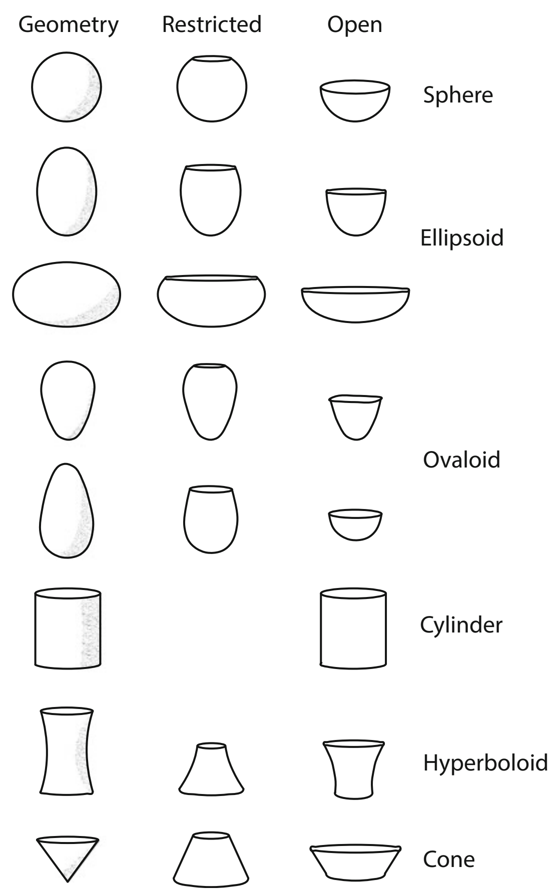


Fig. 12.24 The geometry of overall vessel form can have both functional and stylistic importance. (Y. Salama, after Shepard 1961)

pointed bottom, which served as a sort of third handle. The nearly triangular shape of these vessels made them easier to stack in a ship's hold, the orifice was easy to seal, and the handles and base facilitated carrying and pouring.

The other approach is to look for traces of use on or in the vessels or their fragments (Skibo 1992, 2013). Returning to cooking pots, for example, we might expect them to show signs of thermal spalling on parts of the exterior surface that result from drying wet vessels near an open fire, and from rapid heating during cooking. There may be soot-blackening on the lower part of exterior walls, or oxidation on the bottom, where the pot was in direct contact with flame (Fig. 12.25; Skibo 2013: 84–93). Contact with acidic foods could cause corrosion of interior surfaces, especially when the fabric is calcareous. Frequent washing and scrubbing of pots can cause abrasion on interiors and exteriors, while frequent moving and stacking of pots, and placement of pots on hearth stones or on the ground can cause abrasion and repeated chipping on exteriors and on bases. There could be visible residues of burned food on the interior, or chemical traces of food absorbed into the vessel fabric (Skibo 1992, 2013: 96–99; 161–185).

Today, analysis for absorbed or encrusted chemical residues in pottery is a burgeoning field (Barnard and Eerkens 2017; Gregg et al. 2009; Malainey 2011: 319–331). Unless well sealed, ceramics are porous and vessel walls can contain lipids absorbed from former contents and their degradation products even after many millennia in the ground. Generally, these will reflect an accumulation of contents over a period of time rather than a single use. Analysis requires removal of a small volume of ceramic powder, typically with a high-speed drill, from the sherd, and extracting surviving lipids from the

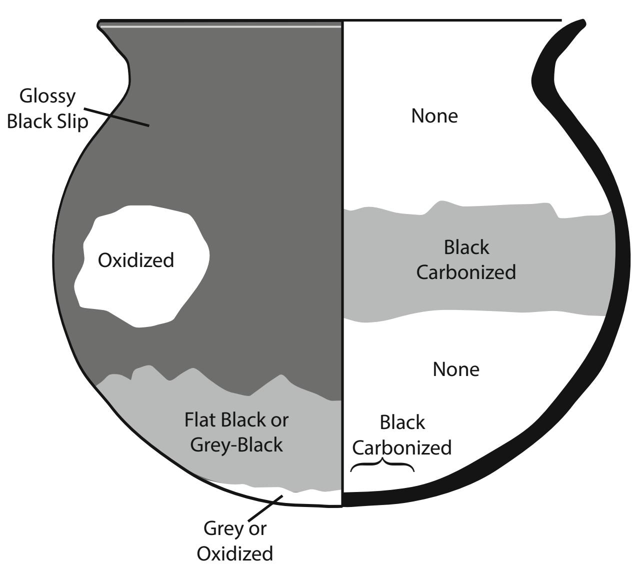


Fig. 12.25 Carbonized deposits and oxidation on a Kalinga rice-cooking pot. (After Skibo 1992: 149)

powder with solvents, sometimes with the assistance of microwaves (Evershed et al. 2002; Gregg and Slater 2010), then a combination of decanting, filtering and centrifuging. Analysis of the extracted compounds by such methods as GC-MS (a combination of gas chromatography and mass spectrometry) allows identification of organic compounds and we can interpret these by comparison with reference fats (see p. 257).

12.6 Stylistic Analysis

Another dimension that archaeologists study in pottery is style, and especially decorative style, which they sometimes take to be evidence of chronological change, social relations, ideology, and even political strategy or resistance (Rice 2015: 402–410). For many decades, ceramic style was one of archaeologists' principal means for defining and refining typological chronologies and regional groups, roles that pottery decoration still plays. The rich variety and potential complexity of decorative and other kinds of style make careful definition of analytical elements, segmentation, differentiation, and orientation rules all the more important. Style is manifested in all kinds of artifacts (see p. 176), but archaeological ceramicists have taken particular interest in it.

Style can occur in all aspects of pottery, not just decoration. Stylistic variation ranges from unconscious variations due to individual potters' motor habits to deliberate symbolic content such as mythological scenes on Greek red-figure vessels. Archaeologists have also attempted to exploit pottery style as evidence for social relationships, such as post-marital residence patterns (Deetz 1965; Hill 1970; Longacre 1970) and communities of practice (e.g., Kohring 2011), and to "decode" the structures of design schemes that underlie decorations from a structuralist perspective (Fig. 12.26; Shanks and Tilley 1987: 160).

Archaeologists differ in how they define and use style (see also pp. 176–177). Some consider it to result from the unique motor habits of individuals, of habits passed on through apprenticeship, or of conscious or sub-conscious social strategies. As with other artifacts, style can also be defined as any kind of variation that does not have substantive effect on the artifact's utilitarian functionality (although it may have considerable impact on social or ideological functions).

Stylistic variation can result from motor habits because no two people will hold or use a tool, such as a paint brush or incising stylus, in exactly the same way, or execute brush stokes, incisions, or impressions in exactly the same order (Fig. 12.27; Hardin 1977; Hill 1977). Prehistorians' use of this approach is similar to that of hand-writing analysts, epigraphers, and art historians, who use small variations in measurements of letters and brush or pen strokes to identify the work of individual writers or artists. For example, Hackett

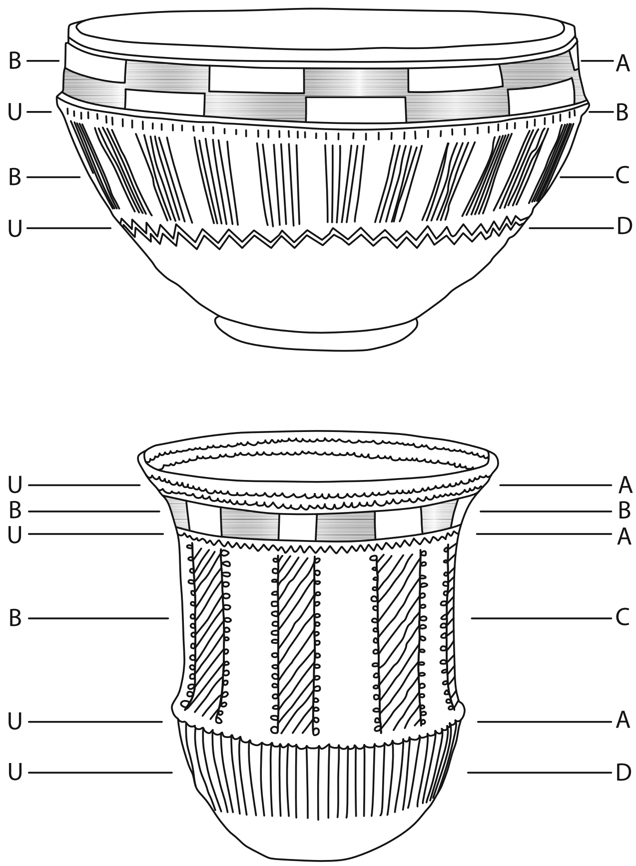


Fig. 12.26 The sequence of distinctive bands of decoration from top to bottom of a Swedish Neolithic vessel can be treated as a sort of code to infer the structure of a design, irrespective of the content of the design elements used in the bands (Y. Salama, after Shanks and Tilley 1987: 160). The letters at left indicated bounded (B, enclosed by a border) and unbounded (U) design elements. Letters at right artibrarily label particular design elements to identify patterns, much like the rhyming scheme of poetry

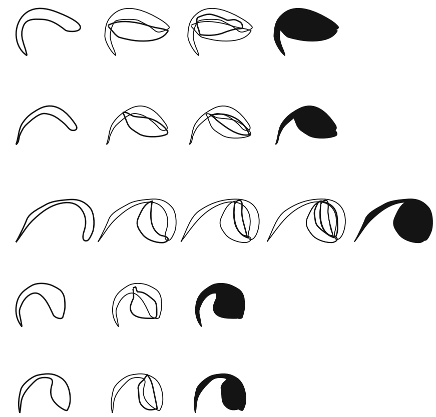


Fig. 12.27 Variation among five individuals in their execution of a painted leaf on pottery from San José, Mexico, with new strokes in heavy lines and final leaf in black. Note variation in the number, sequence, curvature and thickness of brush strokes. (After Hardin 1977: 121)

(1984) uses minute details of brush strokes in his epigraphic analysis of the Deir ‘Alla inscription, an Aramaic document from Jordan that probably dates to the late eighth century B.C. Such micro-analyses are not unique to writing, incising or painting, however. We can make similar kinds of measurements on rim shapes and other pottery features in an attempt to find variations in attributes, or combinations of attributes, that could result from the unique and unconscious or subconscious habits of potters. Apprentice potters are highly unlikely to pick up exactly the same habits as their mentors, even if they may acquire some of them.

Most archaeologists have instead opted to treat style as learned behavior, passed from master to apprentice, or parent to child. It is attractive to assume that style passes on this way because it facilitates inferences of social structure, social interaction, and stylistic changes over time. Classic studies entailed the assumptions that potters were women and learned their craft from their mothers. If these assumptions and some others about pottery disposal were correct, spatial clustering of pottery styles within settlements would be consistent with matrilocal residence pattern (Deetz 1965; Hill 1970; Longacre 1970). Unfortunately, no independent confirmation is available for these assumptions, while a number of ethnographic studies have shown that traditional potters may make deliberate stylistic choices, and do not necessarily learn only from their mothers or mentors (e.g., Hardin 1977: 112–114, 135).

A promising approach exploits small variations in *chaînes opératoires* to identify “communities of practice.” This more fluid conception of community does not require us to assume the gender or cultural affiliation of potters, or even the permanence of their membership in a group, but only how they tended to learn from one another. Rather than focus on unique variations due to individuals’ motor habits, this approach highlights habits that potters probably learned by participating and interacting in a potting community, whether they lived in spatial proximity or not (Canuto and Yaeger 2000; Sinopoli 1991: 120). Potters learned from their interactions, not only with people, but with raw materials and pots.

Case Study

Neolithic Pottery in the Social Fabric of Neolithic Orkney

Andrew Jones (2002) provides an example of using multiple sources of evidence, including petrology and distribution of pottery shapes, volumes and use-wear, to explore social structures and meaning at the Neolithic site of Barnhouse in the Orkney Islands (Fig. 12.28).

Jones classifies the Grooved-ware pottery from the site according to attributes of size or capacity, decoration, temper recipes, and organic residues, some of which we might expect to be related to the use and performance characteristics of the vessels, while others might be more flexible. Turning to just one aspect of his analysis, petrographic examination shows that the pottery exhibits six major fabric groups distinguished by their temper recipes, some of them with rock temper, some with shell, and two with no temper, but there are also variations in the sources of rock used for temper. Jones argues that the fabric groups occur in non-random association with vessel size (Table 12.3), as do the content and structure of decoration.

More interestingly, the fabric groups are spatially patterned (Jones 1997: 192–201; 2002: 126–131). Fabric C occurs in association with structures near the central area of the site, while rock tempers only occur in more peripheral structures. Furthermore, the sources of rock temper vary, with igneous dyke rock 1 only found in structures 5 and 8, dyke rock 5 only in structure 3, and dyke rock 6 only in structure 8. Only structures 3 and 8 have almost all the rock types, but three rock types occur in all the structures. Dykes 2 and 3 are some 2.1 km west of Barnhouse, near the Neolithic tomb of Unstan,

Table 12.3 Prevalence (X) of fabric groups A through E among large (10–35 L), medium (2–8 L) and small (2–3 L) vessels, the latter two categories distinguished largely by wall thickness

	A	B	B1	C	D	E
Large	X	X	X			
Medium	X			X		
Small				X	X	X

After Jones (2002: 122)

Fabrics A through B1 are rock-tempered, C is shell-tempered, and D and E are un-tempered

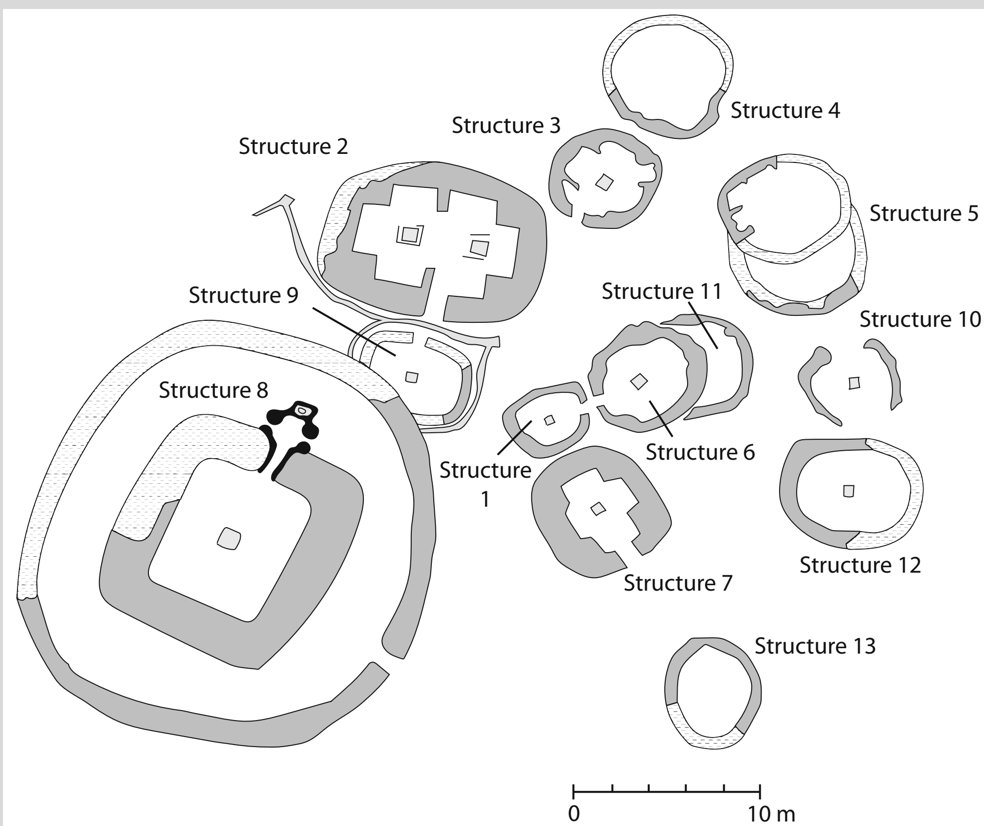


Fig. 12.28 Schematic map of significant Neolithic architecture at the Barnhouse site, Orkney Islands (Y. Salama, after Richards 1990: 309). Structures 2, 3, 4, 5, 8 and 9 are considered “peripheral” to the

(continued)

Table 12.4 Distribution of temper sources for pottery, at dichotomous scale, in four peripheral structures at Barnhouse

Rock type	Structure 2	Structure 3	Structure 5	Structure 8
Sandstone	X	X	X	X
Siltstone	X	X		X
Mudstone	X	X		
Dyke rock 1			X	X
Dyke rock 2	X	X	X	X
Dyke rock 3	X	X	X	X
Dyke rock 4		X		X
Dyke rock 5		X		
Dyke rock 6				X

After Jones (2002: 128)

Tempers from different sources were combined in the pottery found in these structures. Structure 8 belongs to the most recent phase at the site, and mixes most of the rock types found in earlier structures

dykes 4 and 5 are about 3.1 km to the northwest, near the later Neolithic Ring of Bookan, and dyke 6 is some 6.5 km to the northwest, at the far end of the Loch of Harray (Table 12.4).

Jones suggests interpreting these patterns as indicating identification with a place of the dead (dykes 2 and 3, found in all houses), and with a nearby settlement (dykes 4 and 5, found in only two structures). That tempers combined rocks of different sources may itself have been an expression of identity and a metaphor for social relationships within and across communities, while the latest structure, 8, could be interpreted as combining these identities to emphasize community (Jones 2002: 130).

Aside from this specific interpretation, a natural question is whether the alleged pattern in this distribution is statistically significant, or could have arisen just by chance. Unfortunately, presence and absence data do not give us a very full picture, as it could matter whether the ubiquitous rock types, sandstone and dyke rocks 2 and 3, occur in most or all of the pots from each structure, or only some of them, and whether less ubiquitous ones, such as dyke rocks 5 and 6, occur in all pots from one structure, or only some. In some vessels, at least, multiple dyke sources were used to temper individual vessels (Jones 1997: 197; 2002: 130).

Case Study

Quality Assessment and Lot Acceptance Sampling for Chalcolithic Pottery from Jordan

As noted elsewhere in this book, archaeologists need to concern themselves with uncertainty and quality in their data, including ceramic data (Hazenfratz-Marks 2017). In a study of quality assessment for measurements on Chalcolithic pottery from the site of Tubna in Wadi Ziqlab, Jordan, Danielle Cornacchia (2010) carried out a study of inter-observer variation among student volunteers by a combination of control charts and lot acceptance sampling.

Lot acceptance sampling is a method that the US Department of Defense developed to ensure the quality of products in its supply chain, and is an aspect of statistical process control, but the methods are also applicable to archaeology, as in evaluating the quality of archaeological surveys (Almagro-Gorbea et al. 2002). Its key features are to reduce variability through regular inspection of "lots" selected at random from a particular interval of work, to accept or reject a lot on the basis of its meeting or not meeting a specification (tolerance), and to use the results of these assessments to determine whether the frequency of lot inspections should increase, decrease, or stay the same (Table 12.5; MIL-STD-105E 1989). Acceptance sampling plans can involve assessment of attributes on either a continuous interval scale, such as tolerance in some measurement in millimeters, or a dichotomous one (conforming or non-conforming to a standard). There are many other variations on lot acceptance sampling.

Cornacchia adopted a version of acceptance sampling with a single-sampling plan. She modeled non-conforming measurements with the hypergeometric distribution and its binomial approximation (see chap. 8) and used a single random sample of n observations by lab personnel over a fixed period (one week). The standard for comparison was a set of measurements by an expert graduate student, who had analyzed many hundreds of sherds from this same site.[1] The quality measures she assessed were the error rate (see p. 12) for nominal-scale

Table 12.5 Sample sizes required at different verification levels (showing only to level VI) for lot sizes fairly typical in archaeology

	Verification Level (VL)						
Lot Size	VI	V	IV	III	II	I	R
2–170	All	All	80	32	12	5	3
171–288	All	192	80	32	12	6	3
289–544	512	192	80	32	16	8	3

R is relaxed inspection and, for small lot sizes, levels V and greater may require 100% sampling fraction

(continued)

[1] In reality, the expert measurements had already occurred several years previous to the study but were treated as acceptance samples for the purposes of testing the protocol. In effect, the sampling order was reversed.

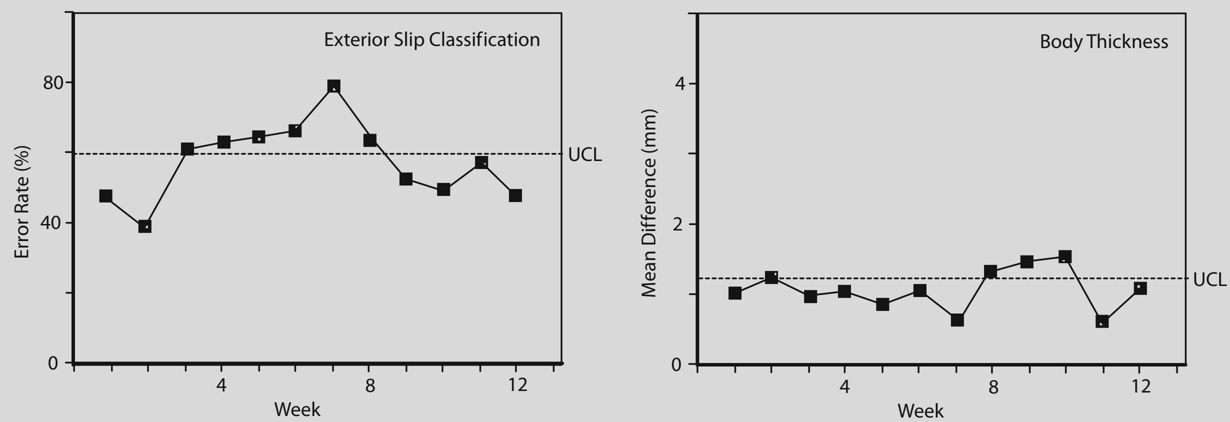


Fig. 12.29 Control charts for error rate in classifying exterior slip (left) and mean difference in measurement of wall thickness (right). Notably the error rate in slip classification is high, and exceeds the Control Limit from weeks 4 to 7. While errors on body thickness are relatively small, they exceed the Control Limit in weeks 8 to 10

data and the mean difference between the expert and all lab personnel for continuous ratio-scale data. To set the frequency of sampling, she followed a revised and more stringent protocol (MIL-STD-1916 1996). Application of the protocol varies, with high verification levels (VL) for critically important variables and lower VL for less critical ones, each associated with a different sample size. "Switching rules" determine the conditions in terms of non-conformance that require shifting from normal to tightened or allow more relaxed sampling regimes. For example, ten consecutive lots must be acceptable at normal inspection before relaxed inspection may be allowed.

The study employed 12 lots of 83 sherds each and initially the low VL of I, in keeping with the small scale of the study and the relatively non-critical nature of the measurements. Consequently, the initial sample size for comparison with the standard was 5 sherds at normal inspection. On each of the 83 sherds in each lot, after a training period and with the aid of an instruction manual, students recorded 36 attributes on standard recording forms. Initially, there were 16 student participants, but the number fell to 7 by the end of the study ("mortality effect," see p. 92).

The control limits for the study were three times the Standard Error of a proportion (for the misclassifications) and three times the SE of the mean (for ratio-scale measurements of differences). Values falling outside 3SE were non-conforming. For illustration purposes, she highlights the results for measurements of interior and exterior slip (nominal scale) and body wall thickness (continuous ratio scale).

Figure 12.29 shows control charts for the results of the acceptance sampling for exterior slip error rates and mean difference in wall thickness. Whenever the estimated errors exceeded the control limits, or there were five consecutive acceptable samples, a change in verification level was required.

12.7 Summary

- Archaeologists have fairly standardized segmentation and differentiation rules for describing pottery vessels
- Archaeologists use a theoretical framework called ceramic ecology or, increasingly, *chaîne opératoire*, to help them understand pottery technology, from acquiring raw materials through construction and firing of vessels, to use, recycling and discard
- They use a variety of macroscopic, microscopic and chemical analyses of pottery attributes to infer the sources of raw materials, the extent of specialization, membership in "communities of practice," and the function and use of pottery vessels
- Style is a major focus of pottery analyses, ranging from the use of decorative features to infer chronology and cultural interactions through the inference of "communities of practice" that learned decorative techniques and motifs from one another
- Technological features of pottery can help us infer even social and ideological aspects of pottery use
- Lot acceptance sampling is one of the tools that archaeologists can use to protect the quality of measurements made on pottery

References Cited

Almagro-Gorbea, M., Benito-López, J. E., & Martín Bravo, A. M. (2002). Statistical quality control in archaeological survey. *Archaeological Prospection*. https://doi.org/10.1002/arp.179.

Applebaum, N., & Applebaum, Y. (2005). The use of medical Computed Tomography (CT) imaging in the study of ceramic and clay archaeological artifacts from the ancient Near East. In M. Uda, G. Demortier, & I. Nakai (Eds.), *X-rays for archaeology* (pp. 231–245). Dordrecht: Springer.

Barnard, H., & Eerkens, J. W. (2017). Assessing vessel fulnction by organic residue analysis. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 625–647). Oxford: Oxford University Press.

Bibby, G. (1969). *Looking for Dilmun*. New York: Mentor Books.

Braekmans, D., & Degryse, P. (2017). Petrography: Optical microscopy. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 233–265). Oxford: Oxford University Press.

Braun, D. P. (1983). Pots as tools. In J. Moore & A. Keene (Eds.), *Archaeological hammers and theories* (pp. 107–134). New York: Academic.

Bronitsky, G., & Hamer, R. (1986). Experiments in ceramic technology: The effects of various tempering materials on impact and thermal-shock resistance. *American Antiquity, 51*, 89–101.

Bunzel, R. L. (1972). *The pueblo potter, a study of creative imagination in primitive art*. New York: Dover Publications.

Callander, J. G. (1937). Report on the pottery. In C. S. T. Calder, (Ed.), A Neolithic double-chambered cairn of the stalled type and later structures on the Calf of Eday, Orkney. *Proceedings of the Society of Antiquaries of Scotland,* 71, 115–154.

Canuto, M. A., & Yaeger, J. (2000). *Archaeology of communities*. London: Routledge.

Cornacchia, D. Y. (2010). *Quality assessment of ceramic attribute data using a combination of inspection by lot acceptance sampling and customized statistical control charts.* M.A. Research Paper. Toronto: Department of Anthropology, University of Toronto.

Courty, M. A., & Roux, V. (1995). Identification of wheel-throwing on the basis of ceramic surface features and microfabrics. *Journal of Archaeological Science, 22*, 17–50.

David, B., McNiven, I. J., Leavesley, M., Barker, B., Mandui, H., Richards, T., et al. (2012). A new ceramic assemblage from Caution Bay, south coast mainland Papua New Guinea: The Linear Shell Edge-Impressed tradition from Bogi 1. *Journal of Pacific Archaeology, 3*, 73–89.

Deetz, J. (1965). *The dynamics of stylistic change in arikara ceramics* (Illinois studies in anthropology 4). Urbana: University of Illinois Press.

Di Angelo, L., Di Stefano, P., & Pane, C. (2018). An automatic method for pottery fragments analysis. *Measurement, 128*, 138–148.

Evershed, R. P., Dudd, S. N., Copley, M. S., & Mukherjee, A. (2002). Identification of animal fats via compound specific ^{13}C values of individual fatty acids: Assessments of results for reference fats and lipid extracts of archaeological pottery vessels. *Documenta Praehistorica, 29*, 73–96.

Franken, H. J. (1974). *In search of the Jericho potters* (North Holland ceramic studies in archaeology 1). Leiden: North Holland Publishing.

Franken, H. J., & Kalsbeek, J. (1975). *Potters of a medieval village in the Jordan Valley*. Leiden: North Holland Publishing.

Gibson, A., & Woods, A. (1997). *Prehistoric pottery for the archaeologist*. London: Leicester University Press.

Gosselain, O. P. (1998). Social and technical identity in a clay crystal ball. In M. T. Stark (Ed.), *The archaeology of social boundaries* (pp. 78–106). Washington: Smithsonian Institution Press.

Gregg, M. W., Banning, E. B., Gibbs, K., & Slater, G. F. (2009). Subsistence practices and pottery use in Neolithic Jordan: Molecular and isotopic evidence. *Journal of Archaeological Science, 36*(4), 937–946.

Gregg, M. W., & Slater, G. F. (2010). A new method for extraction, isolation and transesterification of free fatty acids from archaeological pottery. *Archaeometry, 52*(5), 833–854.

Hackett, J. A. (1984). *The Balaam text from Deir 'Alla*. Chico: Scholars Press.

Hamer, F. (1975). *The potter's dictionary of materials and techniques*. London: Pitman Publishing.

Hardin, M. A. (1977). Individual style in San José pottery painting: The role of deliberate choice. In J. Hill & J. Gunn (Eds.), *The individual in prehistory* (pp. 109–136). New York: Academic.

Hazenfratz-Marks, R. (2017). Evaluating data uncertainty in ceramic analysis. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 48–57). Oxford: Oxford University Press.

Henrickson, E., & McDonald, M. (1983). Ceramic form and function: An ethnographic search and archaeological application. *American Anthropologist, 85*, 630–643.

Hill, J. N. (1970). *Broken K Pueblo: Prehistoric social organization in the American southwest*. Tucson: University of Arizona Press.

Hill, J. N. (1977). Individual variability in ceramics and the study of prehistoric social organization. In J. Hill & J. Gunn (Eds.), *The individual in prehistory* (pp. 55–108). New York: Academic.

Hodges, H. (1964). *Artifacts: An introduction to early materials and technology*. London: J. Baker.

Hunt, A. M. W. (Ed.). (2017). *The Oxford handbook of archaeological ceramic analysis*. Oxford: Oxford University Press.

Hurley, W. M. (1979). *Prehistoric cordage* (Manuals on archaeology 3). Washington: Taraxacum.

Jeffra, C. D. (2015). Experimental approaches to archaeological ceramics: Unifying disparate methodologies with the *chaîne opératoire*. *Archaeological and Anthropological Sciences, 7*, 141–149.

Jizba, Z. V. (1971). Mathematical analysis of grain orientation. In R. E. Carver (Ed.), *Procedures in sedimentary technology* (pp. 313–333). New York: Wiley.

Jones, A. M. (1997). *A biography of ceramics: Food and culture in late Neolithic Orkney*. PhD diss., University of Glasgow.

Jones, A. M. (2002). *Archaeological theory and scientific practice: Topics in contemporary archaeology*. Cambridge: Cambridge University Press.

Kaiser, T., & Lucius, W. (1989). Pottery firing temperatures. In G. Bronitsky (Ed.), *Pottery technology: Ideas and approaches* (pp. 83–94). Boulder: Westview Press.

Karaskik, A., & Smilansky, U. (2008). 3D scanning technology as a standard archaeological tool for pottery analysis: Practice and theory. *Journal of Archaeological Science, 35*, 1148–1168.

Kempe, R. C., & Harvey, A. P. (Eds.). (1983). *The petrology of archaeological artifacts*. Oxford: Clarendon Press.

Kingery, W. D., Bowen, H. K., & Uhlmann, D. R. (1976). *Introduction to ceramics* (2nd ed.). New York: Wiley.

Kohring, S. (2011). Social complexity as a multi-scalar concept: Pottery technologies, 'communities of practice' and the Bell Beaker phenomenon. *Norwegian Archaeological Review, 44*(2), 145–163.

Koul, D. K., Nambi, K. S. V., Gupta, P. K., & Bhat, C. L. (2000). Can quartz remember and manifest its firing temperature through thermoluminescence? *Indian Journal of Pure and Applied Physics, 38*, 243–247.

Longacre, W. A. (1970). *Archaeology and anthropology: A case study* (Anthropological papers of the University of Arizona 17). Tucson: University of Arizona Press.

Lyons, D., Ferguson, J., Harlow, D., & Casey, J. (2018). Marginalized potters and ceramic compositional groups: Neutron activation analysis of contemporary pottery from Tigray, northern highland Ethiopia. *African Archaeological Review, 35*, 567–595.

Maggetti, M. (1982). Phase analysis and its significance for technology and origin. In J. S. Olin & A. D. Franklin (Eds.), *Archaeological ceramics* (pp. 121–133). Washington: Smithsonian Institution Press.

Malainey, M. E. (2011). *A consumer's guide to archaeological science*. New York: Springer.

Mangueira, G. M., Toledo, R., Teixeira, S., & Franco, R. W. A. (2011). A study of the firing temperature of archaeological pottery by X-ray diffraction and electron paramagnetic resonance. *Journal of Physics and Chemistry, of Solids, 72*(2), 92–96.

Martínez-Carrillo, A. L., & Barceló, J. A. (2017). Formal typology of Iberian ceramic vessels by morphometric analysis. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 585–602). Oxford: Oxford University Press.

Matson, F. R. (1965). Ceramic ecology: An approach to the study of the early cultures of the Near East. In F. R. Matson (Ed.), *Ceramics and man* (pp. 202–217). London: Thames and Hudson.

Mayes, P. (1962). The firing of a second pottery kiln of Romano-British type at Boston, Lincolnshire. *Archaeometry, 5*, 80–85.

MIL-STD-105E. (1989). *Sampling procedures and tables for inspection by attributes*. Washington: Department of Defense.

MIL-STD-1916. (1996). *Preferred methods for acceptance of product*. Washington: Department of Defense.

Montana, G. (2017). Ceramic raw materials. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 87–100). Oxford: Oxford University Press.

Nesse, W. D. (1986). *Introduction to optical mineralogy*. Oxford: Oxford University Press.

Pétrequin, P., & Pétrequin, A. M. (2017). *Archéologie en Bourgogne-Franche-Comté: Clairvaux (Jura) et le Néolithique Moyen*. Besançon: DRAC.

Quinn, P. S. (2013). *Ceramic petrography: The interpretation of archaeological pottery and related artefacts in thin section*. Oxford: Archaeopress.

Rasmussen, K. L., De La Fuente, G., Bond, A. D., Mathiesen, K. K., & Vera, S. D. (2012). Pottery firing temperatures: A new method for determining the firing temperature of ceramics and burnt clay. *Journal of Archaeological Science, 39*(6), 1705–1716.

Rice, P. M. (2015). *Pottery analysis: A sourcebook* (2nd ed.). Chicago: University of Chicago.

Richards, C. C. (1990). Postscript: The Late Neolithic settlement complex at Barnhouse Farm, Stenness. In C. Renfrew (Ed.), *The prehistory of Orkney* (pp. 305–316). Edinburgh: Edinburgh University Press.

Roberts, J. P. (1963). Determination of the firing temperature of ancient ceramic by measurement of thermal expansion. *Archaeometry, 6*, 21–25.

Roux, V. (2017). Ceramic manufacture: The *chaîne opératoire* approach. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 101–113). Oxford: Oxford University Press.

Roux, V., & Courty, M. A. (1998). Identification of wheel-fashioning methods: Technological analysis of 4th-3rd millennium BC oriental ceramics. *Journal of Archaeological Science, 25*, 747–763.

Rye, O. S. (1981). *Pottery technology: Principles and reconstruction*. Washington: Taraxacum.

Rye, O. S., & Evans, C. (1976). *Traditional pottery techniques of Pakistan: Field and laboratory studies* (Smithsonian contributions to anthropology 21). Washington: Smithsonian Institution.

Sampson, C. G. (1988). *Stylistic boundaries among mobile hunter-foragers*. Washington, DC: Smithsonian Institution.

Scholfield, J. F. (1948). *Primitive pottery* (Handbook series 3). Capetown: The South African Archaeological Society.

Shanks, M., & Tilley, C. (1987). *Re-constructing archaeology. Theory and practice*. Cambridge: Cambridge University Press.

Shepard, A. O. (1961). *Ceramics for the archaeologist* (Carnegie Institution publications 609). Washington, DC: Carnegie Institution.

Sinopoli, C. M. (1991). *Approaches to archaeological ceramics*. New York: Plenum Press.

Skibo, J. M. (1992). *Pottery function: A use-alteration perspective*. New York: Plenum Press.

Skibo, J. M. (2013). *Understanding pottery function*. New York: Springer.

Smith, M. (1985). Toward an economic interpretation of ceramics: Relating vessel size and shape to use. In B. A. Nelson (Ed.), *Decoding prehistoric ceramics* (pp. 254–309). Carbondale, IL: Southern Illinois University Press.

Tite, M. S. (1969). Determination of the firing temperature of ancient ceramics by measurement of thermal expansion. *Nature, 222*, 81.

Van Gijn, A., & Lammers-Keijsers, Y. (2010). Toolkits for ceramic production: Informal tools and the importance of high power use-wear analysis. *Bulletin de la Société préhistorique française, 107*(4), 755–762.

Vieugué, J. (2015). What were the recycled potsherds used for? Use-wear analysis of Early Neolithic ceramic tools from Bulgaria (6100–5600 cal. BC). *Journal of Archaeological Science, 58*, 89–102.

Waksman, Y. (2017). Provenance studies: Productions and compositional groups. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 148–161). Oxford: Oxford University Press.

Whitbread, I. K. (2017). Fabric description of archaeological ceramics. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 200–216). Oxford: Oxford University Press.

Wilczek, J., Monna, F., Jebrane, A., Chazal, C. L., Navarro, N., Couette, S., et al. (2018). Computer-assisted orientation and drawing of archaeological pottery. *Journal of Computing and Cultural Heritage, 11*(4), 1–17.

Zhu, J., Zhang, Y., Wang, T., Zhao, C. H., Yu, J. C., Glascock, M. D., et al. (2013). Determining the firing temperature of low-fired ancient pottery: An example from the Donghulin site, Beijing, China. *Archaeometry, 56*(4), 562–572.

13 Metal Artifacts

The first smiths had discovered that a hard and intractable reddish substance, copper, became malleable and plastic on heating. . . . It can be shaped by casting into forms the old materials could never assume The change in the properties of copper by heat is really very startling

(Childe 1930: 3–4).

Metal must surely have been a very surprising material to those who first discovered or encountered it, even if it took many centuries for them to appreciate its qualities more fully (Smith 1975). Even though its production depends on technologies related to other materials, such as ground stone and pottery, its properties are radically different from those of most of the stones with which prehistoric people were most familiar. Ancient metal artifacts have consequently been of great interest to archaeologists, but also to metallurgists. In fact, a great deal of the research on ancient metal artifacts falls into a sub-discipline called **archaeometallurgy**, which focuses on such topics as the provenance of ores, chemical and isotopic composition and microstructures of the metals, and the technologies used to make alloys and artifacts (Franklin 1999; Killick 2015; Killick and Fenn 2012; La Niece et al. 2007; Tylecote 1992; Roberts and Thornton 2014; Rothenberg 1978). However, archaeometallurgy also concerns the non-metallic aspects of these technologies, such as stone picks and hammers for mining and breaking up ore, stone or ceramic crucibles and *tuyères*, fuel for furnaces and forges, and stone or ceramic molds. This young field can be so specialized that it may seem isolated from mainstream archaeology, or risk projecting modern industrialist concerns onto the past, yet more recent researchers have attempted to bring to it more archaeological concerns, including social and aesthetic factors, and even magical beliefs, in the interpretation of metal artifacts and metallurgy (Bray and Pollard 2012; Budd and Taylor 1995; Hosler 1995; Ottaway 2002).

Metal and metallurgy-related artifacts are extremely diverse. They include slag and crucibles from metal production (Bayley and Rehren 2007), ingots, weapons, woodworking tools, vessels, coins, jewelry, horse trappings, locks and other hardware, and even cut-marks on bones (Greenfield 2013). Metal artifacts can be small components of furniture or vehicles, such as nails, studs, hinges or drawer pulls in chests, cabinetry and carriages, or huge components of bridges and buildings. They can be rare but are sometimes extremely numerous; the site of a single historic barn, for example, could easily contain more than 5000 nails. No wonder they were historically shipped in barrels.

Some of the earliest metal artifacts are cutting tools that probably served as either real or symbolic weapons, notably axes and knives, spearheads, and swords. Only somewhat later did metal almost completely replace stone for more mundane cutting tools, such as sickles, chisels, and scrapers.

Another common type of metal artifact is the metal vessel. Archaeologists describe and illustrate the forms of metal vessels in much the same way as pottery (see Chaps. 12 and 21), but the fact that their *chaînes opératoires* are very different requires very different approaches to documenting their material and technology. Metal vessels also often had social or symbolic roles quite different from those of more common vessels.

13.1 *Chaînes Opératoires* for Metal Artifacts

Variable as metal tools are, their *chaînes opératoires* also vary considerably, while having some fundamental aspects in common. An alternative model to the *chaîne opératoire* that some archaeometallurgists employ, called the archaeometellurgical cycle (Fig. 13.1, Ottaway 1994), highlights these more common elements. Notable among them is the ease with which smiths can recycle metal tools by melting them down. This has major implications for the survival rates of metal artifacts, especially for ones made from scarce or valuable metals.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_13

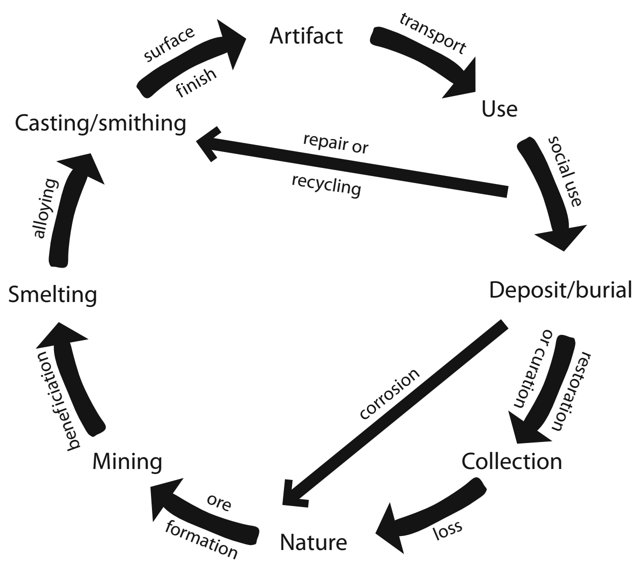


Fig. 13.1 The "archaeometellurgical cycle," somewhat similar to the *chaîne opératoire*. (Modified from Ottaway 1994). Note that the cycle for ferrous technology would differ somewhat because iron can be welded and repaired in ways that copper cannot

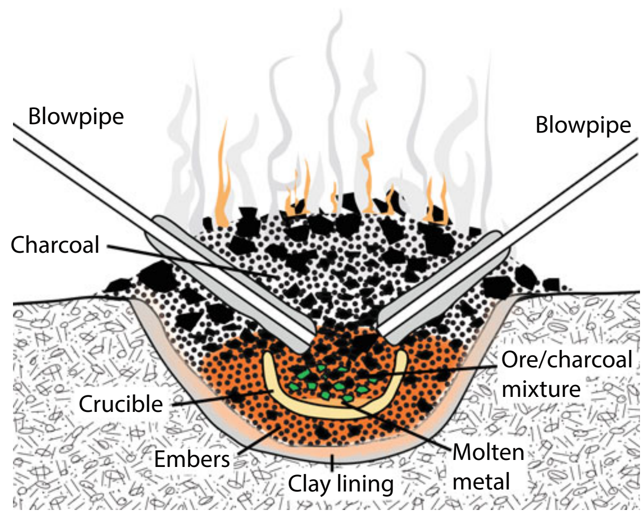


Fig. 13.2 Cross-section of a simple copper-smelting bowl furnace with a clay lining, crucible and blowpipes

13.1.1 Raw Materials, Metals and Alloys

Some of the earliest metallic artifacts were made of native copper, silver or gold, simply hammered from rather pure, naturally occurring nuggets or veins of metal without altering their chemical characteristics (Halsey 1996; Wayman 1985). Native copper is by far the most notable of these technologies. It can be shaped by hammering as long as one anneals (heats) it periodically to prevent it from becoming brittle. Among the more famous examples are the large, shield-like coppers of North America's Northwest Coast, although surviving examples are of European trade copper, and it is not clear that such large artifacts could have been made from native copper nuggets (Jopling 1989: 72–73; but see Deans 1885). Clearer prehistoric examples are the Archaic-period copper artifacts of the Lake Superior Region (Bebber and Eren 2018; Martin 1990), Hopewellian and Mississippian artifacts (Chastain et al. 2011; Ehrhardt 2009), and more recent artifacts of Alaska and the Yukon (Cooper 2011, 2012).

Except when found as nuggets in rivers, native metal had to be extracted from its parent rock either by prying it out or by heating the rock with fire (fire-setting) before pouring cold water on it so that sudden thermal changes shattered it (Jopling 1989: 74).

In addition, well before people knew how to smelt iron, they made use of meteoric iron (Hulme 1937; Zimmer 1917). Iron from meteors is characterized by high concentrations of nickel, sometimes with some cobalt. Although it is malleable, it is difficult to work, but its celestial origins probably enhanced its desirability in antiquity. Some noteworthy examples of its use in antiquity are Predynastic beads found at Gerzah, Egypt (Hulme 1937: 181), and a dagger from Tutankhamun's tomb (Comelli et al. 2016). Artifacts of meteoric iron are also known from the New World (Halsey 1996: 3; Zimmer 1917).

In several parts of the world, however, people independently realized that heating certain kinds of rocks to a high temperature could release a pool of molten metal that, when cool, could be hammered into useful tools in the same way as native metals. In the case of copper, the minimum temperature is 1089 °C, too high to maintain very long in an open fire, while it is also necessary to have a reducing (oxygen-starved) atmosphere to remove oxygen and hydrogen from the ore. For these and other reasons, people needed to be skilled in their use of pyrotechnology — most likely through making lime and pottery — before they could develop metallurgy.

Important prehistoric copper mines included those in Bulgaria, Cyprus, Hubei in China, eastern Egypt and Sudan, southwestern Ireland, Laos, the southern Levant, western Mexico, Peru, Serbia, and Wales (Ben-Yosef 2018a, b; Hosler and Macfarlane 1996; Tucci et al. 2014). Reaching temperatures that exceed 1200 °C to smelt copper from ores, such as the sulfide ores chalcopyrite ($CuFeS_2$), chalcocite (Cu_2S), and covellite (CuS), carbonate ore malachite ($CuCO_3 \cdot Cu(OH)_2$), or oxide ores like cuprite (Cu_2O), requires a furnace.

A very simple furnace is merely a hole in the ground, sometimes lined with clay, into which the smiths placed a mixture of charcoal and crushed ore, and sometimes inserted several blowpipes for oxygenating the fire (Fig. 13.2). Often, archaeometallurgists conduct experiments or ethnoarchaeology to help them understand how such furnaces might have operated and what their challenges were (e.g., Timberlake 2007).

The smelting process varies considerably, so what follows is just one simplified version for copper. In the first stage of the extraction process, roasting at relatively low temperature, in the presence of oxygen (with wind, blowpipes or bellows) helps decompose copper-bearing minerals and drive carbon, hydrogen, and sulfur into the atmosphere in the forms of CO_2, water and SO_2. Especially for sulfide ores, this is often done in a dedicated roasting pit, prior to use of the furnace. The second stage, reduction, requires a higher temperature and a reducing (oxygen-starved) environment in the presence of carbon from the charcoal. This is easier to accomplish in a furnace set in a deep hole, as the overlying charcoal impedes the entry of oxygen, but is also possible by placing a piece of turf or a pile of dung on top of the fire. First carbon monoxide and later carbon dioxide bond with the oxygen in the CuO, and CO_2 is expelled, leaving the copper. In some cases, limestone mixed with the ore acts as a flux to remove other impurities and concentrate them in slag. Iron-rich ores may require addition of silicate to help create slag. After the furnace had cooled, the smiths would find small prills or a sort of ingot of metal pooled at the bottom of the pit, sometimes underneath slag, a glassy mixture of metal oxides and sulfides, and silicon dioxide.

To a degree, the characteristics of metal that results from these smelts depend on the minerals in the ores. Although the first metal artifacts were often made from relatively pure metals with only traces of other elements, later ones were usually made from alloys — materials made from smelted ores and containing more than one metallic element. These require a complex *chaîne opératoire* of mining ore, processing the ore to extract desirable elements, melting and mixing these to make metal alloys, and then casting the metal into ingots, jewelry or tools, sometimes with further steps such as annealing and hammering to make a finished product (Craddock 1995). Intentional alloying of copper with tin was one of the most important of these developments. However, early alloys could result from the use of ores that already contained multiple metals. For example, *fahlerz* ores (mostly the copper mineral, tennantite) naturally contain small amounts of arsenic, zinc, silver, and iron.

The fact that metallic ore sources vary in their chemical and isotopic characteristics presents opportunities to distinguish the ore sources used for particular artifacts and groups of artifacts. However, a number of problems typically intervene to make this very difficult except under special circumstances. One is that metal artifacts' composition may tell us more about the maker's metallurgical skills and preferences than about the metal sources. Another is that overlapping distributions of variability in ore sources can lead to error (Villa 2009). The most serious problem is that metal was often both valuable and easily recyclable, so that a typical artifact contains material blended from metals of multiple sources.

Among the earlier successes of archaeometric sourcing of metals was the detection of sources for early Greek silver coinage (Gale 1979; Gentner et al. 1978). By focusing on the earliest coins, researchers benefit from the probability that recycling had less impact than would be the case for coins made several centuries later, although this does not mitigate the possibility of exploiting multiple sources, or of recycling jewelry or other silver artifacts. The silver for these coins came from lead ores, and Greek metallurgy depended on an inefficient method called cupellation to extract silver from the lead; this makes it possible to source the traces of lead left in the coins by their distinctive ratios in lead isotope*s* (^{204}Pb, ^{206}Pb, ^{207}Pb, ^{208}Pb; De Muynck et al. 2008*)*. Lead ore deposits of different ages vary in these ratios because some of the isotopes are created through the radioactive decay of uranium.

The first step in these analyses is to analyze ores from known sources, typically using Inductively Coupled Plasma spectroscopy (ICP). Scatterplots of the ratios of the isotopes ^{208}Pb and ^{207}Pb to ^{206}Pb show distinct clustering of the ores from Laurion, the source closest to Athens, from the island of Thasos, and from Ayos Sostis on the island of Siphnos, while more distant sources in Sardinia and Spain are more widely separated (Fig. 13.3). Ores from Macedonia and Thrace plot in the region between Laurion and Siphnos, close to the cluster for Thasos, and there can be overlap between some sources that makes source identification less certain (e.g., Gale et al. 1997; Stos-Gale et al. 1996).

In one of the early studies (Gentner et al. 1978), data from Archaic Greek coins found in the Asyut hoard in Egypt show lead-isotope ratios that associate most of them clearly with one of these sources. As expected, all of the Athenian coins have silver from the mines at Laurion, while coins of Aegina and Corinth sometimes show silver from Laurion, sometimes from Ayos Sostis, and sometimes occupy the field between those clusters, perhaps indicating a blend of metals from the two sources. Some coins of Thasos plot near the Ayos Sostis field but show other compositional differences that hint at a different source.

13.1.2 Alloys

Pure copper is too soft to hold an edge very well but, when alloyed with other metallic substances, it becomes harder (Craddock 1995, 2001). Consequently, among the first intentional alloys was bronze, an alloy of copper with tin, antimony, arsenic, or their combination, which is more resistant to wear than unalloyed copper. As soon as early metallurgists realized that copper containing one of these elements was easier to melt, and harder and more resilient once cooled, they began to seek out materials they could blend with the copper. In many parts of the world, the choice was first arsenic, but

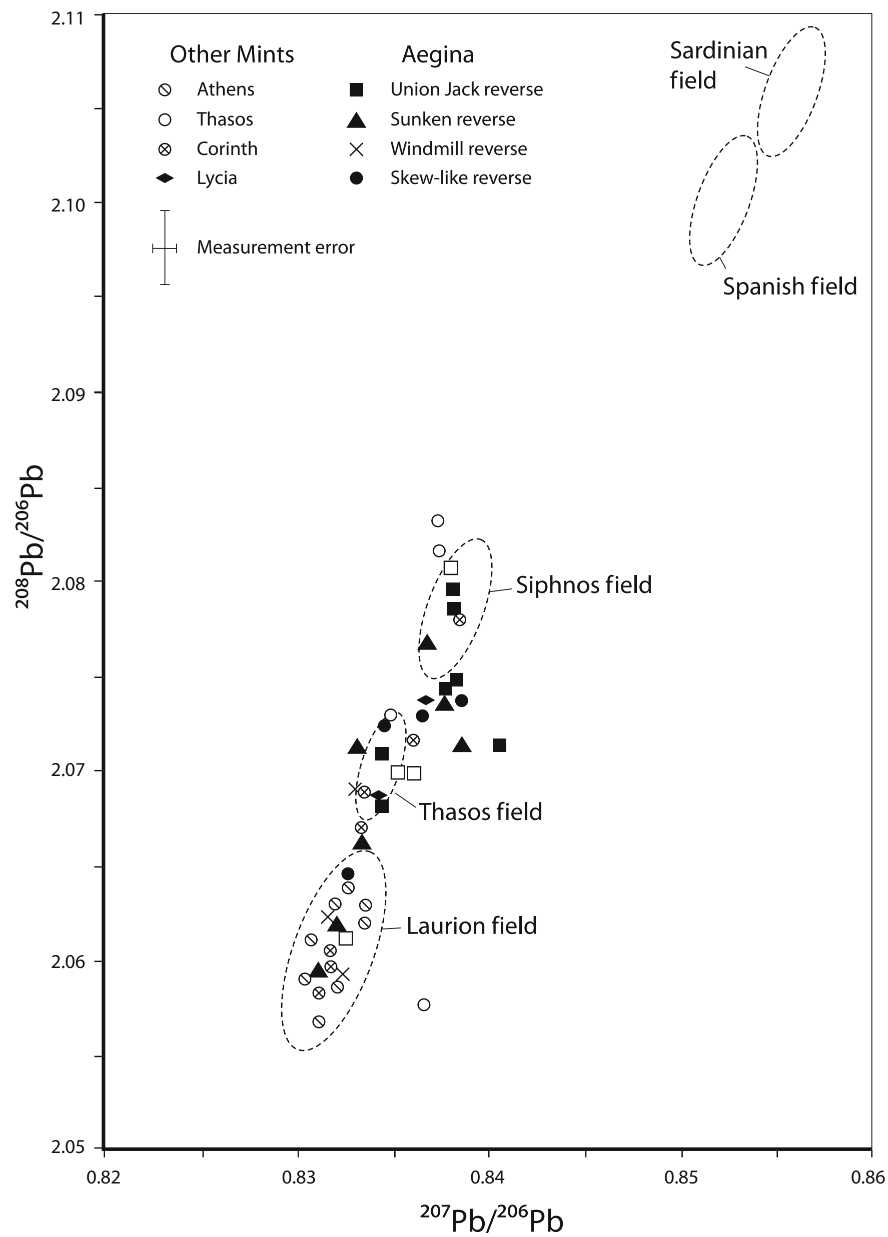


Fig. 13.3 Scatterplot of lead-isotope ratios of silver sources and early Greek silver coins. (Modified from Gentner et al. 1978: 279). The ellipses show the regions associated with ores and slags from Laurion, Thasos, and Ayos Sostis (Siphnos), with Sardinian and Spanish sources at far upper right. For Aeginetan coins, the different symbols distinguish different reverse designs. Locations for Athenian coins are only approximate, but all fall within the Laurion field

the less toxic tin later became more common. They could either add one of these to a melt, or they could smelt different kinds of ore in the same furnace, to produce the alloy from the beginning. Eventually, they became quite skilled at controlling the ratios of the various metals in their alloys. Aside from bronze, some of the more historically important alloys include gold-silver (**electrum** is a naturally occurring gold-silver alloy), silver-copper, brass (copper-zinc-tin), pewter (tin-copper-lead) and steel, an alloy of iron and 1 or 2% carbon. As already noted, meteoric iron is essentially an iron-nickel alloy.

As metal has usually been valuable and can easily be re-melted, recycling also tends to be a very important aspect of metallic *chaînes opératoires*. Resharpening can maintain the edges of metal tools for a time but, once this has worn the tool too much, one can re-melt the valuable metal to make new tools rather than discarding it.

Repeated recycling of metal can actually be detectable. Pollard and Bray (2014) point out that some constituents of copper alloys, such as arsenic and antimony, become depleted the longer the metal is in a molten state, and therefore also the more times it is melted. Recycling can of course also result in the mixing of metals of different sources or different alloying recipes. They suggest that these compositional changes allow us to work out something like a biography of bronze artifacts. Focusing on artifacts made from copper mined on Ross Island in southwestern Ireland, they find that that mine's distinctive alloy of copper, arsenic, antimony and silver was used to make Early Bronze Age daggers and axes found throughout Ireland and Britain. The chemical variations in the alloy are consistent with the hypothesis that artifacts found farther from the Ross Island source had experienced more episodes of remelting than ones near the source, having passed through a longer chain of such events as the metal moved into more distant regions. Often, the latest re-casting resulted in an artifact that fit locally prevalent designs rather than any of the shapes found in southwestern Ireland.

Plating and surface enrichment are methods used to make the surface of metal artifacts made from baser alloys look like more valuable ones, typically silver and gold. Plating is a technology that fuses a thin layer of gold or silver onto the substrate alloy, and in antiquity the most common method to accomplish this involved the use of mercury as a sort of "glue" to attach gold leaf (cold mercury gilding), or applying an amalgam of gold or silver with mercury to the surface and then heating the artifact to about 300-350 °C to volatize the mercury and leave the precious metal behind. Burnishing was typically then necessary to smooth the surface and remove bubbles and pores. Surface enrichment, by contrast, involves treating an artifact made from an alloy that contains a relatively small amount of gold or silver, so the more valuable metal becomes concentrated at the surface. For example, Roman mints of the second and third centuries AD apparently tried to hide debasements of coin metal and make many of their coins look like silver. One way they may have done this was heating the coin blanks so that copper oxides formed on the surface, and then pickling them in acids or salts to remove copper and leave surficial silver behind before striking the blanks between coining dies (Cope 1972: 267). When surface enrichment is used to create a gold surface, it is called depletion gilding, commonly used for Mesoamerican and Andean "tumbaga" (copper-gold) artifacts.

Just as archaeometallurgists have investigated ore sources, they have also invested a great deal of research into the nature of alloys. While sourcing has tended to depend more on isotopes, especially lead isotopes, these studies use a suite of analytical methods, such as X-ray fluorescence (XRF) and proton-induced X-ray emission (PIXE), to determine the concentrations of various elements in the metal. For more information on how these methods work, there are many publications (e.g., Agarwal 1991; Aitken 1974; Goffer 2007; Pollard et al. 2014; Price and Burton 2012; Roberts and Thornton 2014; Shugar and Mass 2012).

Many of these methods characterize elemental compositions of materials by an electromagnetic **spectrum**. The typical output is a graph with the energy of detected photons on the x-axis and the number of such photons per "bin" on the y-axis (intensity), much like a histogram with very small intervals (Fig. 13.4). We can associate the positions of peaks along the x-axis with particular elements by careful examination of the associated energies, and the areas of the peaks above the "background" noise are related to the abundances of those elements in the material. However, there is no one-to-one correspondence between peak area and abundance — we need to do careful calibrations

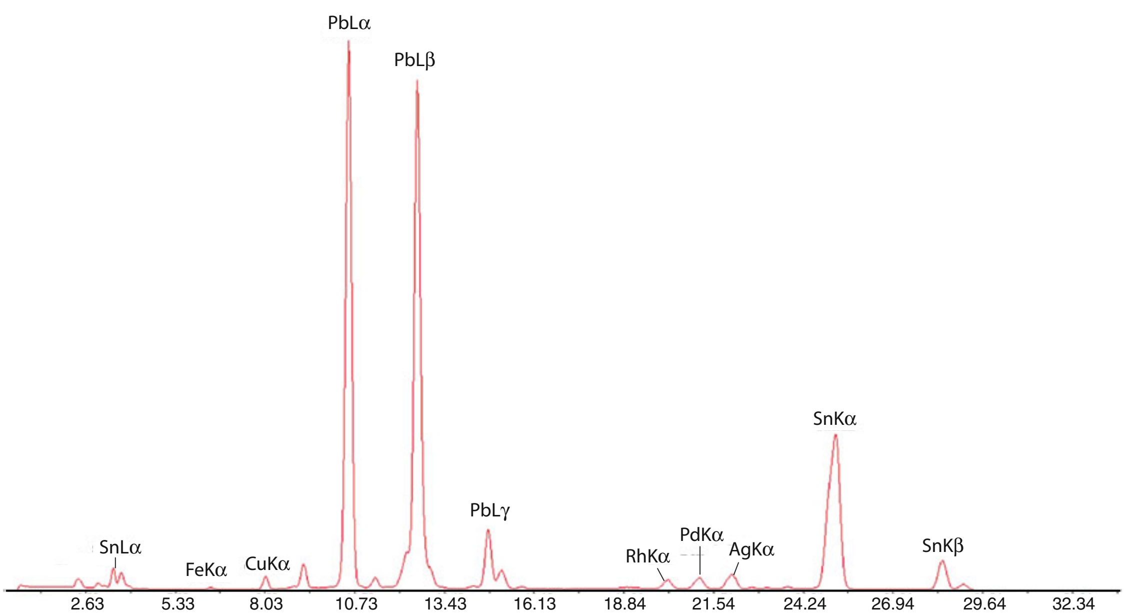


Fig. 13.4 Example of an X-ray spectrum from pXRF analsysis of a nineteenth-century pewter medal. Note the prominent K_{α} and K_{β} peaks of tin (Sn), L_{α}, L_{β}, and L_{γ} peaks of lead (Pb), and smaller peaks that might indicate very small amounts of iron (Fe), copper (Cu), rhodium (Rh), palladium (Pd) and silver (Ag). Units are keV (kiloelectron volts)

based on measurements on standards (alloys with known composition) to relate peak areas to abundance — and even the identification of each peak with some element is not always obvious. For example, in the case of XRF, the energies of photons caused by fluorescence of an element results in more than one peak, so if we see a large peak around 8.046 keV (kiloelectron volts) that we think comes from the K_α fluorescence of copper, we need to check to make sure that there is a smaller peak where we would expect to find photons from the K_β fluorescence, at 8.905 keV. We also need to be aware that some peaks in the spectrum may be spurious ones, in the sense that they result from secondary fluorescence within the material, from overlaps between peaks of different elements, from materials in the instrument itself, or from bremsstrahlung radiation that comes from deceleration of electrons in the material.

13.1.3 Shaping Metal

Shaping both native copper and meteoric iron involved hammering with a hammerstone on an anvil and hammering later became important for making sheet metal and many tools and weapons made from smelted metals. Because hammering copper not only spreads the metal out, but also hardens it, it is necessary to anneal it periodically with heat to soften it and allow further hammering (Deans 1885). The result was often sheet metal of non-uniform thickness that could then be cut with chisels or rolled to be further shaped, sometimes with further hammering and annealing to produce knife blades or daggers. The Tlingit on the Northwest Coast also knew how to attach copper sheets with rivets to make larger artifacts (Jopling 1989: 76). Some metals, notably iron, need to be hot during hammering, while bronze needs to be cold-hammered to prevent shattering. Riveting involves inserting soft metal cylinders into overlapping holes that pierce the sheets, and then hammering the ends of the cylinders to widen and flatten them so that they join the sheets firmly.

As smelting metal already involved the high temperatures necessary to make molten metal, casting early became an important method for shaping metal tools and weapons. Open casting in a simple mold soon led, in some regions, to lost-wax casting and then complex molds made from several pieces. Some examples of these technologies follow.

13.1.4 Casting of Bronze Cutting Tools

Bronze can be used for a wide array of tools, including vessels like bowls, jewelry, and buckles and horse trappings. The first truly metallic artifacts were often jewelry or symbolic items (Smith 1975) and, in the New World, the symbolic, decorative and ritual uses of metal were dominant right up to the fifteenth century (Hörz and Kallfass 1998; Killick and Fenn 2012: 561; Tushingham et al. 1979). In much of the Old World, however, metal fairly soon became important for making cutting tools and weapons, although it took some time to replace stone cutting tools.

The *chaîne opératoire* of a bronze axe, for example, began with identifying sources of copper and tin, or copper ores that already contained small amounts of tin, arsenic, or antimony. Smiths would extract the metal from these ores using a furnace, as described above.

The resulting copper or bronze could be used to make tools directly by casting or annealing and hammering, or the smiths could cast it into ingots for shipping to a production site or for use in trade. Famous examples of ingots are the "oxhide ingots" of the Bronze Age Mediterranean, hundreds of which were found in the Cape Gelidonya shipwreck (Fig. 13.5; Bass 1967; Muhly et al. 1977; Lo Schiavo et al. 2009).

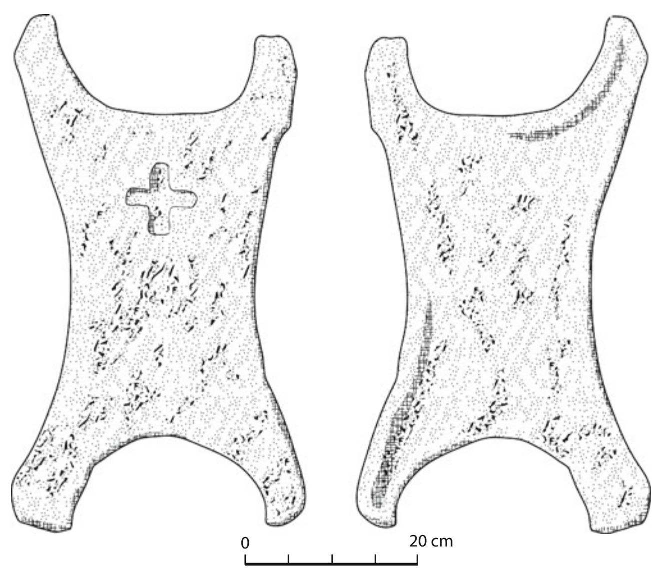


Fig. 13.5 "Oxhide" ingot from the Cape Gelidonya shipwreck site. (After Bass 1967: 56)

Like most early ingots, early axes were cast in open molds of stone, clay, or possibly sand (Fig. 13.6) and then allowed to cool. Subsequent hammering (cold-working) could harden the edge to make it more durable, while grinding the edge would make it very sharp (Heeb 2014). Similar methods, but with much more cold-working, may have been used to manufacture knives and swords (e.g., Sapiro and Webler 2016; Wadsworth 2015), although the exposure of the upper surface of the metal to oxygen during casting of blades was not desirable. Copper axes circulated far from the places where miners extracted the ore. Lead-isotope analyses show that the axe of Ötzi the "ice man," for example, came from southern Tuscany but it was found with Ötzi's body many hundreds of kilometers away in the Alps (Artioli et al. 2017).

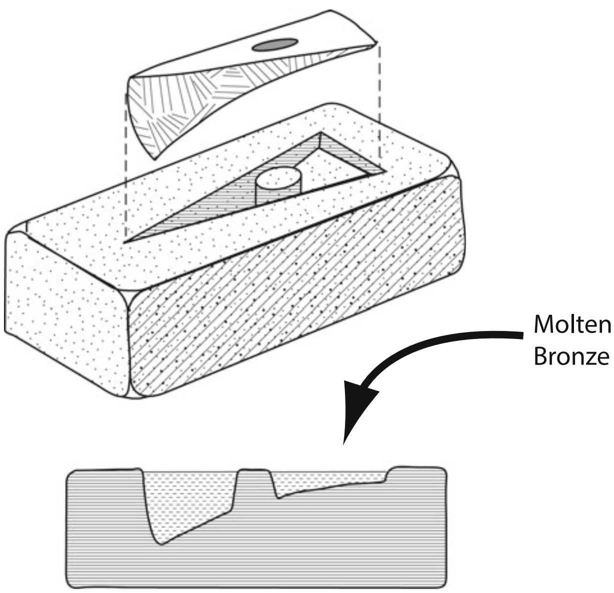


Fig. 13.6 Open casting of a bronze axe with shaft-hole, with cross-section through mold at bottom. (Modified after Hodges 1964: 71)

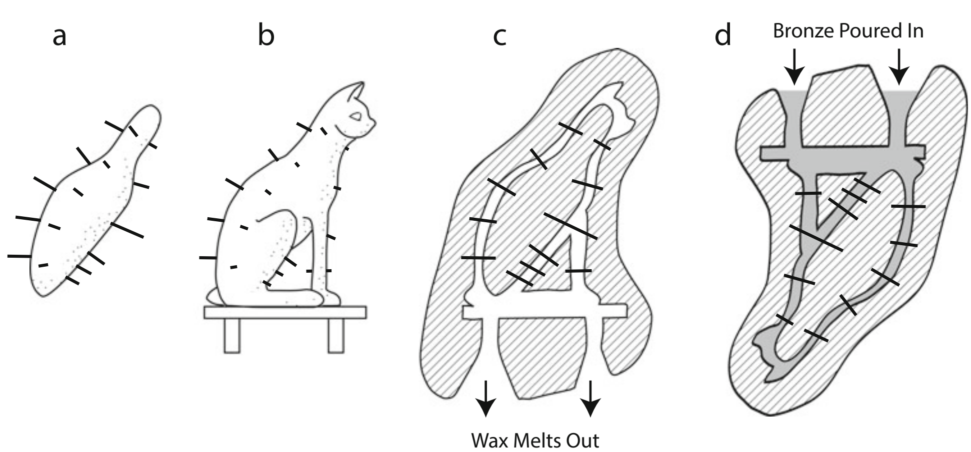
Two-piece molds of moist sand (in wooden frames) or clay later allowed more complex shapes and restricted exposure to oxygen during casts.

13.1.5 Lost-Wax Casting

Ancient metallurgists determined that they could cast more complex shapes by carving a model in wax and surrounding it with clay except for wax "sprues" that connected the interior wax to the exterior of the mold. Often, to save metal, the smith would use a clay core or "investment" separated from the outer mold with pins or "chaplets" that maintain the spacing that would later determine the thickness of the metal casting. Heating this assembly would cause the wax to melt and run out of a "gate" left by a sprue, leaving a hollow ceramic mold and a pool of wax that could be reused. They would then pour molten metal into the still-hot mold through the gate while air in the mold escaped through vents, also left by sprues. After it cooled, the smiths would break the mold open to reveal the desired artifact, which they could fine-tune by removing projections where the sprues had been and any other imperfections, and polish or burnish the surface (Fig. 13.7).

A Chalcolithic (ca. 3600 BC) hoard from the Nahal Mishmar cave, west of the Dead Sea, provides many excellent examples of this technology (Moorey 1988; Tadmor et al. 1995). Among other copper artifacts, the hoard included several vessels, more than 240 pear-shaped copper mace-heads, 90 complex "scepters" or "standards," and ten "crowns." Most of these items had been made with the lost-wax technique, and some of the mace-heads contained the ceramic investment or core that is a remnant of this process. While some of the artifacts were of rather pure copper, most of the complex ones had arsenic content of 1 to 12%, or antimony content of 1 to 25%, possibly the result of smelting copper and arsenic or antimony ores together.

Fig. 13.7 Lost-wax casting of a bronze cat, (**a**) clay core with pins or "chaplets," (**b**) wax figure molded over the core, leaving chaplets protruding, (**c**) figure encased in clay and then baked to allow wax to melt out, and (**d**) molten bronze poured into the "gates." After cooling, the bronze figure is broken out of the clay, the chaplets and gates tooled off, and the surface smoothed and polished. (After Hodges 1964: 72)



Case Study

Ethnoarchaeology of African Ironworking

In their attempts to understand earlier ironworking technologies in Africa, archaeologists have often turned to ethnoarchaeological studies of traditional ironworking. In one such study, Len Pole (2010) interviewed old participants in ironworking in the Volta region of Ghana.

The Togo Hills there are rich in iron ore and the region has been a center of iron production at least since the Mawu settled there in the mid-nineteenth century.

In the 1970s, building a furnace involved four or five men digging a pit into a flattened area in the treeless brow of a hill some distance away from settlement, about 120 cm in diameter and 45 cm deep. After making several ritual deposits (see also Schmidt 1996), they built a foundation at the base of the pit from anthill clay to support the shaft walls, which were built up with an internal diameter about 40 cm, and leaving a hole at the base of the shaft on the side facing downslope, where a gulley was later excavated to drain off the slag (Fig. 13.8). For the next two weeks, the shaft base rested and hardened, except for burnishing its interior with smooth stones to prevent cracks from forming.

Then, the builders added another layer of clay, about 12 cm thick, put three sticks into the wall that, after removal, would leave air inlets and, after some disagreement among the participants about its proper size and shape, enlarged the front hole to a diameter of about 25 cm. They also disagreed about the appropriate thickness of the walls and other aspects of the smelter's design, finally adding a layer some 10–18 cm thick weekly for 8 more weeks. They also blocked the front hole with a plug of clay and pieces of old kilns, leaving only a couple of smaller holes as vents. The total labor requirement was about 50 person-hours.

Once the smelter itself was complete, the men built a platform much like the flat roof of a house. They set four strong poles into the ground to act as corner pillars, laid horizontal beams across the front and back pairs of pillars, and narrower poles across the beams at right angles. They covered these with wooden shingles and about 10 cm of clay, with the chimney of the smelter projecting about 15 cm above this, and put a low parapet around the platform edges, with rain gutters at intervals. This construction protected the smelter from rain but also provided a useful platform from which to charge the furnace from above. A sheet of corrugated iron served to cover the top of the chimney except when in use, to keep out rain. In some accounts, the inside of the smelter received a thin coat of clay or its floor was covered with sand.

The ironworkers charged the smelter with about 80 kg of charcoal of a preferred tree and then added about 15 kg of ore broken into egg-sized pieces followed by another 10 or 12 kg of charcoal, filling the smelter to within 50 cm of the top of the shaft. Then they added a small amount of burning charcoal to start the fire, and left it to burn,

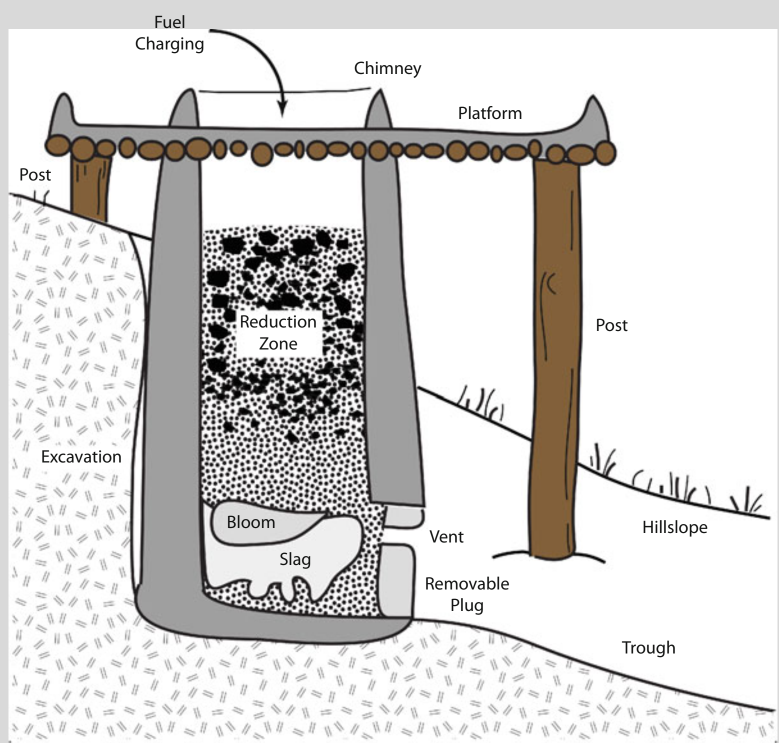


Fig. 13.8 An iron-smelting furnace like those built at Akpafu Todzi, Ghana

(continued)

checking it occasionally. After some 22 hours, they broke open the clay plug blocking the front hole, exposing slag. This particular firing that Pole describes was not successful, as it did not reach a high enough temperature to produce molten iron, in part because they neglected to put a fire at the bottom of the shaft before charging it. This prevented a good updraft of hot gases and the ignition of the carbon monoxide that needs to comingle with the ore for the reduction of iron.

Had this been a successful smelt, the ironworkers would have extracted a **bloom**, a porous mass of slag and iron called sponge iron. A blacksmith would need to work this bloom by heating and hammering to produce wrought iron, and Mawu blacksmiths would use this to fashion iron hammers and other tools, or trade it for cowry shells.

While the obvious aspect of iron metallurgy was the production of usable metal and tools, it is important not to ignore its symbolic and spiritual aspects. In the case of the Mawu, ore had to be kept far away from the presence, or even gaze, of children, menstruating women, and anyone who had recently had sexual intercourse. A libation to the ancestors preceded firing of the smelter, the blacksmith's anvil was considered sacred, and men involved in forging hammers had to avoid quarrels and sexual intercourse before the work (Pole 2010: 59–60, 66). More generally, many metal-making cultures associate metallurgical processes with ritual, taboos, magic, and danger.

13.1.6 Chaînes Opératoires of Metal Vessels

Most metal vessels and vessel-like artifacts, such as helmets, were constructed by hammering sheet metal made malleable by annealing, and sometimes later decorated by punching or incision. Two basic methods for this involved hammering the sheet metal while it rested on a small, convex anvil, or hammering it while it rested in a concave, bowl-like anvil (Fig. 13.9). These methods leave signs in the form of flat, convex or concave facets on the vessel exterior, and concave indentations on the interior surface.

However, casting remained a very important method for making bronze vessels in China (Fig. 13.10). Some of these vessels have very complex shapes, such as animals and boxes with three or four legs, and their manufacture required equally complex multi-part molds (Chase 1991; Fairbank 1962; Loehr 1968; Zhanwei 2008).

As the Chinese drinking vessels contain significant amounts of lead, which improves the flow of molten metal

Fig. 13.10 Replica of a bronze rectangular ding (*fangding*), or cauldron, of the Shang dynasty (ca. 1600–1000 BC)

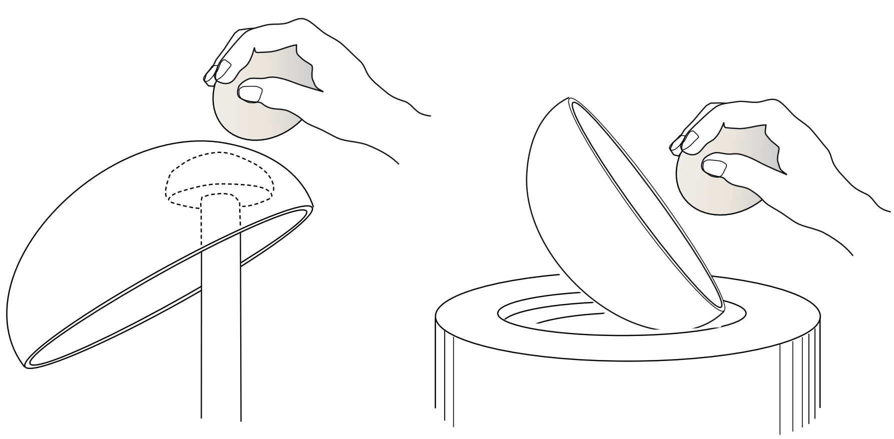


Fig. 13.9 Two methods for forming vessels from sheet metal, by hammering over a convex anvil (left) and hammering into a convex basin. (Right; after Hodges 1964)

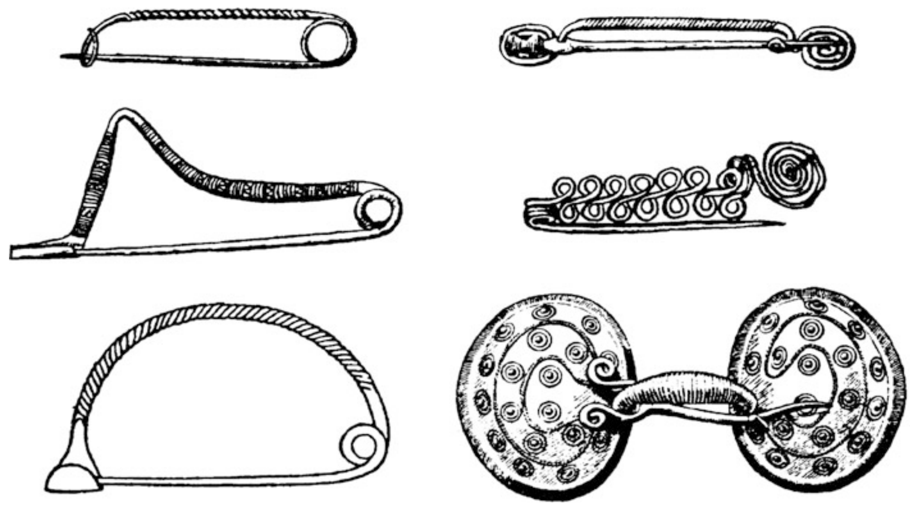


Fig. 13.11 A selection of *fibulae* from the Danube region. (After Childe 1930)

into the molds, using them to drink wine entailed risk of both copper and lead poisoning. Woolf et al. (2010) found that leaving wine in replica Shang Dynasty vessels, even for just 1 day, yielded lead concentrations in the wine of more than 100 mg per liter, although the leaching of lead into the wine would vary with acidity, interior surface area, and other factors.

13.1.7 Jewelry

Pins, brooches, rings, earrings, bracelets and safety pins have a long history of archaeological interest, as these items of personal adornment, often associated with individuals in graves, exhibit tremendous variability that makes them valuable clues to chronology, identity, and socio-economic status.

When application of statistical methods to the grouping and classification of artifacts was in its infancy, *fibulae* (decorative safety pins for fastening cloaks) were among the first subjects of study. Already, many archaeologists, such as V. Gordon Childe (1929: 432–433), had noticed that the styles of *fibulae* from Bronze Age Europe were chronologically sensitive (Fig. 13.11). Mathematicians who collaborated with archaeologists in the 1960s latched onto these as potentially good subjects for "automated" sorting of artifacts in ways that might constitute a chronological seriation (see Chaps. 3 and 18; Doran and Hodson 1975: 218–237). Not surprisingly, what attributes of these complex artifacts to measure and how best to measure them are not obvious, and could have considerable impact on the results.

13.2 Conservation Aspects

Metal, and especially iron, artifacts are particularly vulnerable to corrosion, and often require more than the usual safeguards during handling and storage (see Chap. 9). It is best to avoid handling any metal artifact with bare hands, as the oils and acids in skin become sites for corrosion. Nitrile gloves prevent this. Copper and bronze artifacts should not be kept in containers made of polyvinyl acetate (PVA) or that contain plasticizers, while iron artifacts require storage in environments of low relative humidity and inert containers that will not give off acidic vapours. Use of silica gel in the storage containers and some means of monitoring relative humidity is particularly important for iron artifacts (Shearman 1990).

13.3 Metal as Wealth

Ancient people probably valued metal for a variety of reasons, including not just its practical properties, but its rarity, color, reflectivity when polished, and even the sounds it produces when struck (Killick and Fenn 2012: 562). Consequently, metallic objects have often featured as symbols of wealth and standards of value. Even before the invention of coinage, metal objects and especially standardized ingots probably circulated as a form of currency in some places, eventually contributing to the development of market economies, while later hoards often contain mixtures of coins, metal jewelry and scrap metal. Not surprisingly, the

Case Study

Neolithic Copper Beads

Well before there were coins, it is likely that certain valuable or scarce items served as varieties of currency or "primitive money," perhaps including certain kinds of sea shells and precious metals. Ring-shaped gold and electrum ingots that may have been a sort of currency are known as early as the Chalcolithic (ca. 4000 BC) in Israel (Gopher and Tsuk 1991).

A very interesting early case of possible "primitive money" or currency consists of copper beads from the Neolithic Cortaillod culture (ca. 4000–3500 BC) in the Lake District of western Switzerland. At one Cortaillod site, Burgäschisee-Süd, archaeologists found two strings of copper beads that someone had carefully deposited just outside the north wall of a house. The beads were strung on cords, fortunately preserved by the wet conditions, knotted at their two ends to prevent loss of the beads. String I had 18 relatively narrow but heavy beads (median of 13.3 g), while the other had 36 broader but lighter beads (median of 4.25 g). In both cases, the beads were made from bent strips of copper, with traces of arsenic, antimony, silver, and sometimes bismuth or lead, cut from five bars, all apparently from the same source but at least three different melts (Fig. 13.12). Although the masses of individual beads were not standardized, two beads on string II had had small blobs of metal added to them to increase their mass, and string I had half as many beads while having twice the total mass of the beads on string II (Ottaway and Strahm 1975). This strongly suggests that these strings of copper beads were a form of currency, possibly constituting a standard of value for other commodities, while also clearly demonstrating numeracy (possibly based on 6 or 12) and knowledge of beam balances.

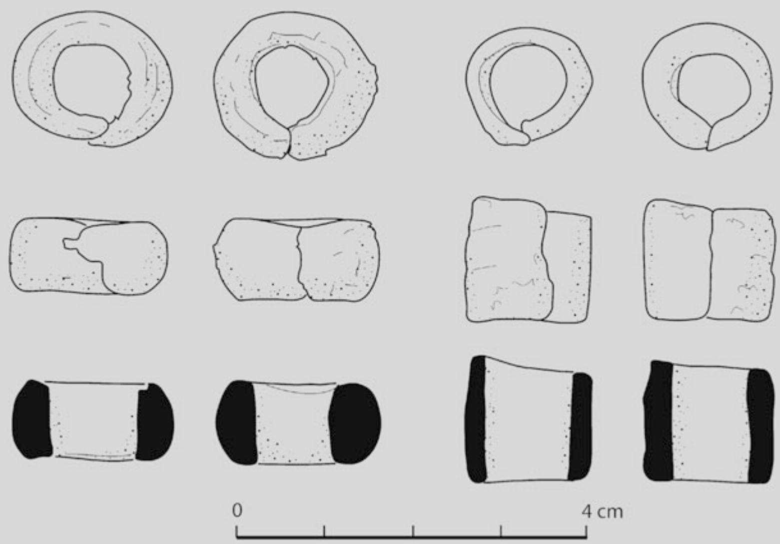


Fig. 13.12 A selection of Neolithic copper beads from two strings found at Seeberg, Burgaschisee-Süd. Beads from string I are heavy and have cross-sections that are rounded on the outside and flat inside. Those from string II are lighter and made from flatter bars. (After Ottaway and Strahm 1975: 312)

search for valuable metals and incentive to control their sources and trade routes have been important stimulants of political change, warfare and imperialism (Killick and Fenn 2012: 559).

13.3.1 Coins

Coins constitute a very special class of metal artifact because, not only have they usually contained an intrinsically valued mass of metal, they, like standard weights, also embody a symbolic value in a very clear way (Kenoyer 2012; Rahmstorf 2012). In most parts of the world that used coins, ancient and mediaeval coin manufacture involved striking a disk of metal, or "planchet," by placing it on an obverse metal die set into an anvil, placing a punch with the reverse die over it, and then hammering the punch sharply to impress the designs engraved into both dies onto the metal (Fig. 13.13; Laing 1969: 6–12). Until fairly recently, however, coins in most parts of eastern Asia were made by

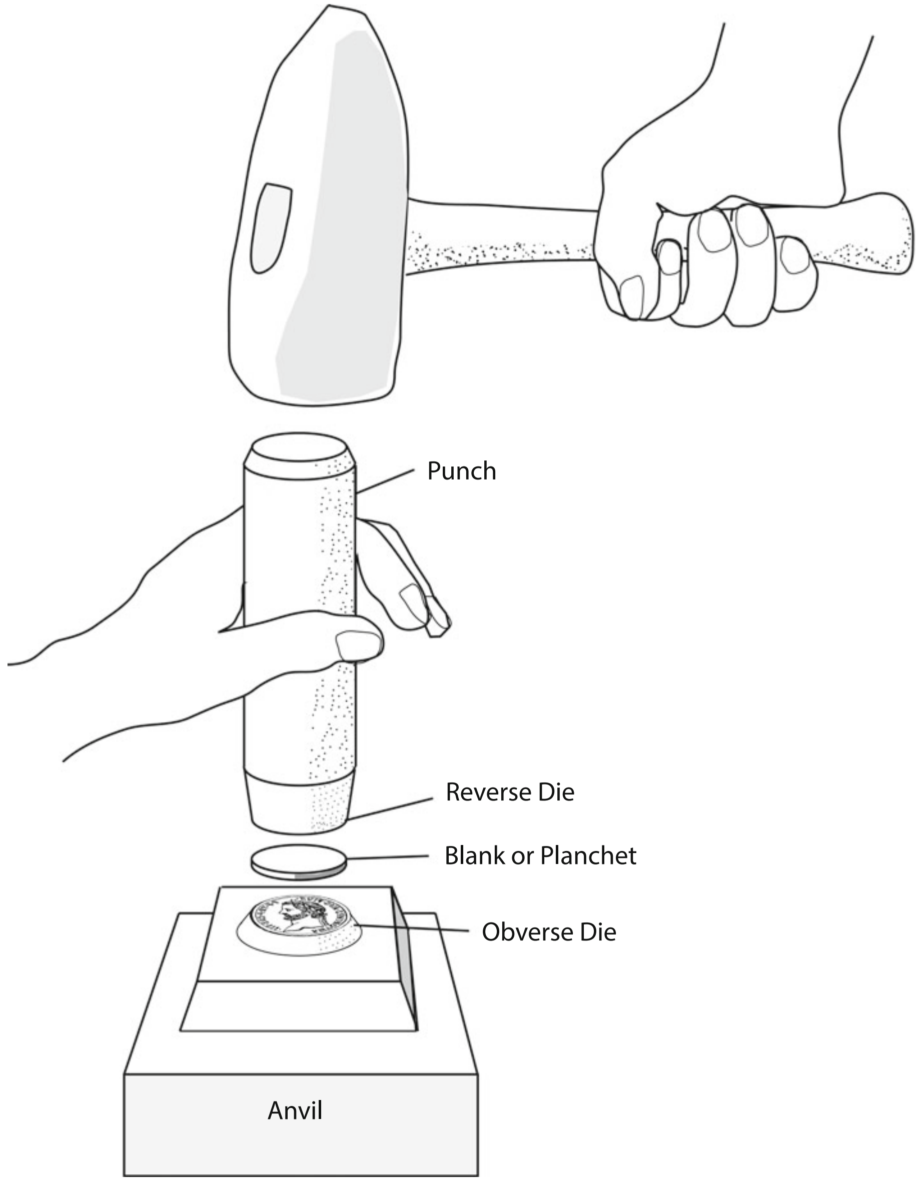


Fig. 13.13 Until the sixteenth century AD, coining in most of the Old World west of China involved hammering a heated metal disk, called a planchet, between two engraved steel dies. The planchets could be cut or punched from sheet metal, sawn from a cylindrical bar, or cast in molds

casting in two-part molds. Because early coins'value depended on the intrinsic value of the metal in them, we can infer ancient monetary systems from the distribution of mass in coins, although we must keep in mind that corrosion and leaching may have altered the mass of some of them (Fig. 13.14).

Because of their key role in economic systems, the distribution of coin types can tell us a lot about ancient economic behavior. For example, the discovery of many Roman silver coins in India helps confirm textual evidence that Rome, especially under Augustus and Tiberius, had a trade imbalance with South Asia, its source for spices, carnelian, and some other luxuries (Laing 1969: 136–138). It is even possible, using probability theory, a method analogous to the Krantz and Peterson indices (see pp. 114–116), and differences between the numbers of obverse and reverse dies to estimate the number of coins that were struck for each issue (Esty 1986; McGovern 1980).

However, coins not only inform us about ancient economies (Casey 1986; Casey and Reece 1974; Kemmers and Myrberg 2011), or even chronology (Lockyear 2012). Because coin issues were often the prerogative of a monarch or a state, and states also varied the mass and fineness of their coins as acts of monetary policy, they provide unusually relevant evidence of ancient political territories (see Case Study: Iron Age Territories in England).

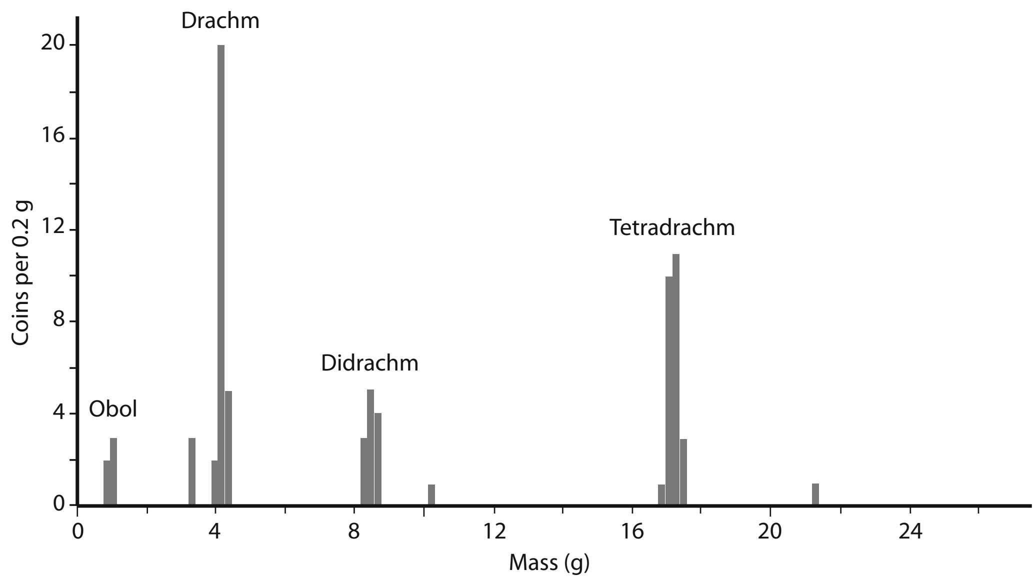


Fig. 13.14 Histogram of the masses of Athenian silver coins from the fifth century B.C. (data from Starr 1970: 23–38). Notice that regularity in the spacing of the peaks of the distribution allow us to distinguish *oboloi* (1/6 drachm), *drachmai*, *didrachmai* (2 drachms), and *tetradrachmai* (4 drachms), even though there is some variation due to minting errors, wear, and post-depositional corrosion

13.3.2 Hoard Analysis

One very helpful aspect of coins is that they are often very well dated. Even though we sometimes need to be suspicious of dates appearing on coins and tokens — they can be anachronistic or even intentionally misleading[1] — they are among the very few artifacts that archaeologists encounter regularly whose manufacture we can often date to a single year. Once in circulation, however, coins can continue in use for many decades, so the date on a coin is not necessarily very closely associated with the date of the deposit in which we find it. In such instances, they provide a *terminus post quem*, or earliest possible date, for the deposit (see Chap. 20).

One very revealing type of deposit is a coin hoard (Casey 1986: 51–67). Hoards are accumulations of wealth, often but not always consisting of coins and bits of valuable scrap metal, and sometimes built up over many years. In ancient times in the Old World, burying a jar filled with coins under the floor or in the courtyard of one's house was a common way to accumulate savings and keep wealth secure from robbers or invaders. Sometimes, for one reason or another, the hoard's owner never retrieved it, leaving it for some later archaeologist or treasure-seeker to discover.

Hoards provide a great opportunity to study the variety of coinage that was in circulation in a particular locality over the period of the hoard's accumulation, and also to give a sense of the use-life of coins. The former requires some analytical caution, however, as hoarders were often selective about the kinds of coins they included in their savings, favoring high-value silver coins, for example, over cheap bronze ones. Thus, they are usually not random samples of the population of coins in circulation, which we would expect to contain mostly recent coins with the frequency of dated coins rapidly decreasing with age (Collis 1974: 178–182). Hoards also provide insights into the value of individual coin finds for dating the deposits in which they are found. Finding a coin dated to the equivalent of 52 B.C. in no way dates the associated deposit to that year, only a *terminus post quem* (see p. 328), the earliest possible date for the deposit, but our knowledge of nearby hoards can help us evaluate the probability that the deposit could be as late as, say, A.D. 100. How long were coins current before they were discontinued or, more likely, melted down to exchange for more current coins? This is a matter of estimating the average use-life of the coin type.

As shown in Fig. 13.17, the distribution of dated coins in a hoard provides clues to the date that the hoard was "closed,"

[1] Dates on coins can sometimes refer to some historical event, rather than the date of issue, while tokens (non-governmental or private coin-like currency) were sometimes intentionally misdated to evade anti-counterfeiting laws, as was the case with halfpenny tokens of Upper Canada dated 1820 but actually struck in the 1830s. Still others — especially ones much in demand for foreign trade — retained the same date for many decades, notably 1780-dated silver thalers of Empress Maria Theresa of Austria, which were still being minted at least as recently as the 1960s.

Case Study

Iron Age Territories in England

Around the time of Julius Caesar's invasion of England, several different Gallic "tribes" issued coins in their territories. Plotting the distribution areas of such coins' find spots on maps allows comparison of their distributions with other spatial tools for estimating the territories of the main settlements at that time, such as Thiessen polygons, which mark boundaries halfway between neighboring towns (Hodder and Orton 1976: 79–80).

To establish the boundaries of the coin distributions, Hodder and Orton imposed a grid on a map of the smoothed densities of the coins. They assumed a boundary between two neighboring coinage areas at a grid edge when the densities of find spots for coins of two different sources were the same on both sides. This does not take into account that the volume of coin production and density of coin finds can vary from region to region, or that the boundaries (much as with histograms) vary with the size of grid unit used, which somewhat weakens its validity. However, it was a useful approximation.

With these caveats, these tribal coins appear to have circulated mainly within the territories of their issuers, and the fairly close correspondence of the coin distributions to the territories that the Thiessen polygons estimate reinforces the plausibility of these reconstructed territories (Fig. 13.15). However, not all artifact distributions show such clear boundary effects (Kimes et al. 1982).

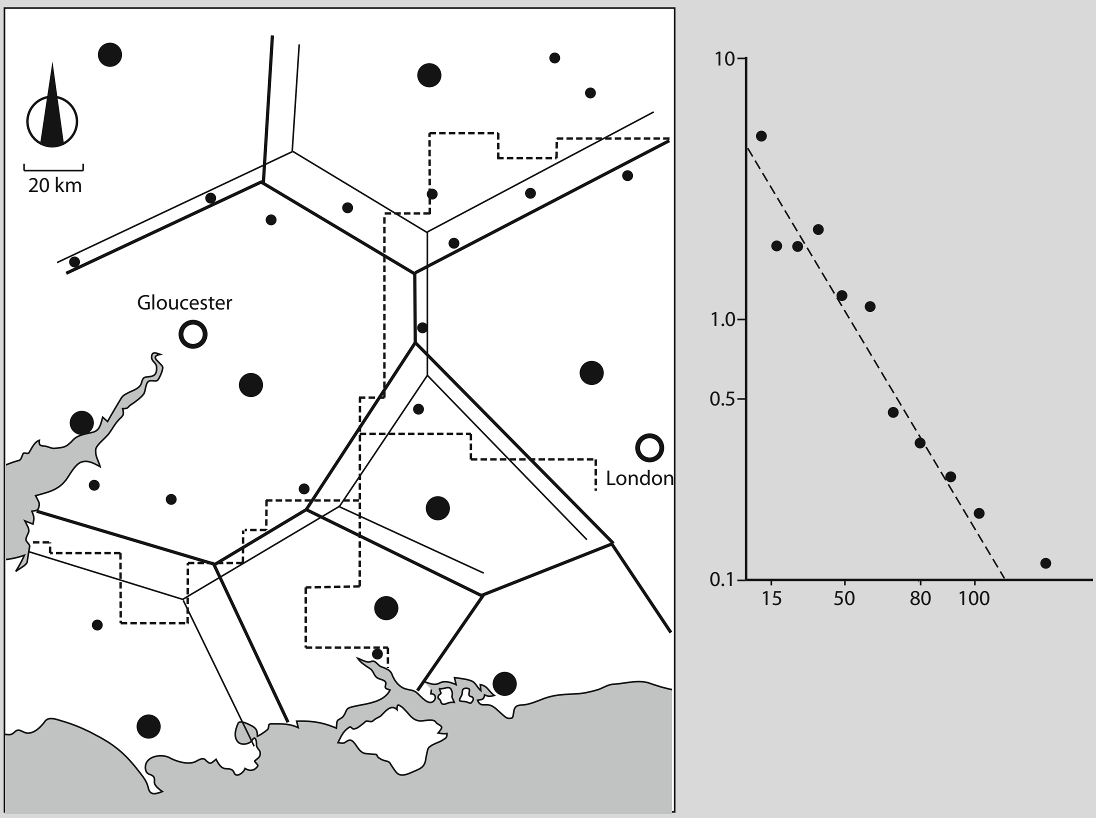


Fig. 13.15 Distribution of the find-spots of late Iron Age Gallic silver coins in southern England (left) along with Thiessen polygons (heavy lines) defined by the mid-distances between cantonal capitals (large filled circles), and weighted Thiessen polygons (thin lines), adjusted for the sizes of the capitals. Small filled circles represent smaller walled towns. The fact that the higher densities of coins issued at each of the capitals tend to be found within their respective Thiessen polygons confirms that these are reasonably close to the actual territorial boundaries of the first century BC (after Hodder and Orton 1976: 79). At right, regression of the density of Dobunnic Iron Age coins against distance from their minting-place. (Note logarithmic scales, after Hodder and Orton 1976: 111)

Case Study

Uneven Circulation of Roman Coins in the Western Empire

Hodder and Reece (1977) used modeling and Trend-Surface Analysis to reveal patterns in the temporal and spatial distributions of Roman coins in Europe that are explicable in terms of Rome's trade and military commitments. Trend-Surface Analysis is something like non-linear regression, but in three dimensions instead of just two, fitting a smoothed "surface" to the data for the frequency of coins at particular locations on the map that expresses the average or trend over space. As in any regression, we need to pay attention to the fit of the data to the model.

The interesting aspects, however, are the locations and sizes of the residuals — positive and negative departures from the predicted trend (Fig. 13.16). Over different time periods, the monetary policies of the emperors Trajan and Severus show impacts in the ratios of small-, middle- and high-value coins. Trajan's policies, for example, appear to have depressed the proportion of middle-value *sestertii*. On the spatial side, positive residuals near the Rhine, dating to the three phases from AD 180 to 238, emphasize that the Roman frontier was showing more marked increases in the proportion of *sestertii* than the trend surfaces predict; these may occur because the soldiers stationed on the frontier required increasing commitments of money for their pay. By contrast, northern Italy in the later periods shows mainly negative residuals, possibly because of lower rates of inflation, while the exceptional positive residuals occur in coastal cities where we would expect heightened levels of commerce.

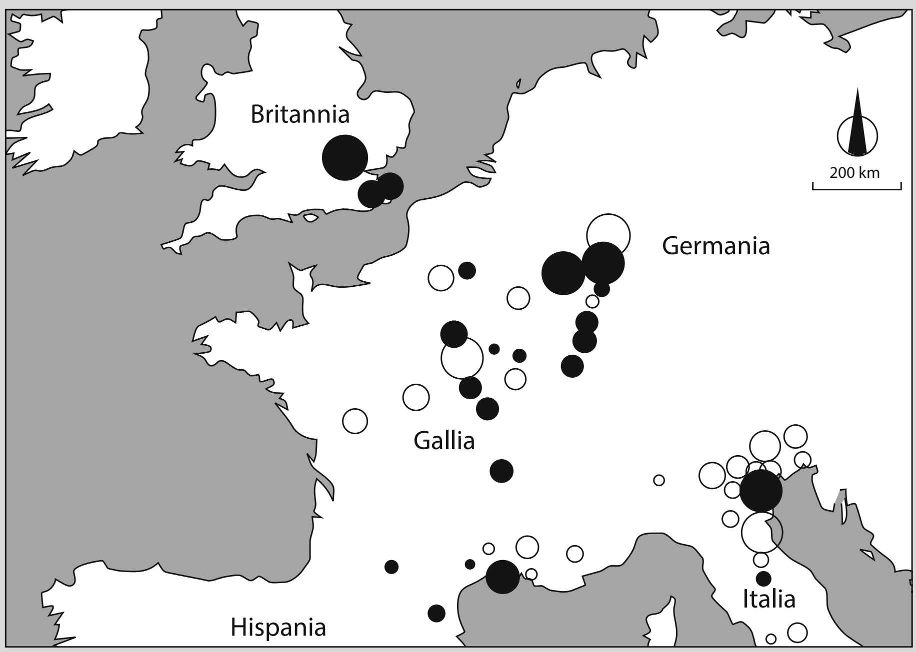


Fig. 13.16 Trend-surface analysis of the distributions of Roman low-value bronze, middle-value *orichalcum* (brass), and high-value silver coins of AD 192–222 in western Europe, with the sizes of circles indicating the size of residuals from the trend surface (black, positive, and white, negative residuals). Note that positive residuals are common near the Rhine and in major ports, such as Masillia (Marseilles), while negative residuals cluster in northern Italy. (After Hodder and Reece 1977: 13)

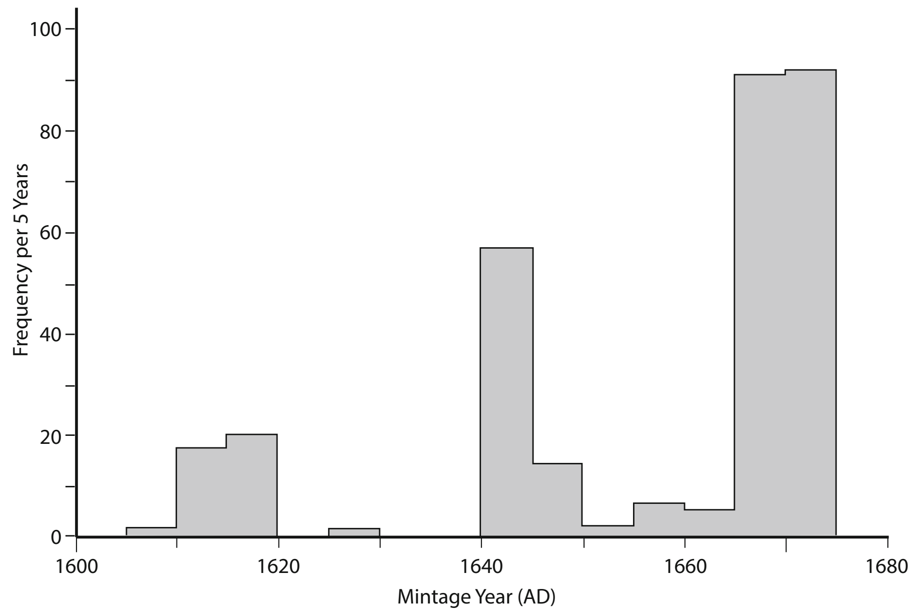


Fig. 13.17 Frequency distribution of dated coins in a hoard from Grønnegade, Rudkøbing, Denmark (Märcher 2012). The latest coins in the hoard are three dating to 1675, suggesting that the hoard was "closed" that year, while 1645–1673 appears to be the main period of accumulation. The older coins in the hoard might indicate that their use-lives extended at least until the time of initial accumulation, but the trimodal distribution suggests that this is not a single savings hoard, but had three distinct periods of accumulation over more than one generation

the likely period of accumulation, and the length of time that some of the older coins in the hoard may have circulated. Typically, the distribution peaks just a few years before the date of the most recent coin in the hoard, and the most recent coins most likely indicate the last year that the owner added anything to the hoard. Moving left of the peak, the distribution may be fairly steady for a period of years that likely corresponds with the period of accumulation, and then drops off to the left, indicating that coins were less and less likely to be included in the hoard the older they were. The oldest coins in the hoard are ones that were nearing the end of their use life at the time that the hoard began to accumulate. Some hoards can be exceptions to this pattern, but that in itself is informative; unusual spikes in the frequency of fairly old coins may be because they were intrinsically more valuable, for example by having a higher silver content, or that the hoard incorporates an older hoard, perhaps inherited.

13.4 Quality in Archaeometallurgical Analyses

While scientific methods for analyzing metallic artifacts are many and potentially very enlightening, there are risks that such analyses can provide inconsistent results or have no clear purpose, while inconsistency among analysts poses particular problems for "Big Data" analyses (pp. 55–56; Ben-Yosef 2018a; Bevan 2012; Pearce 2016). The problems of reliability and validity, in particular, require close attention to the quality of measurements. Among the known problems with data quality, we might include:

Lack of a clear research question: Sometimes it is tempting to analyze the composition of metal artifacts just because we can, whether we are archaeologists or analytical scientists. Without a clear research question, however, it is impossible to select the most appropriate techniques to answer it (Pearce 2016: 47).

Use of inappropriate methods: Often in conjunction with the previous problem, archaeologists or their physical-science colleagues sometimes employ a particular method just because it happens to be available at their institution, or because they happen to have that particular expertise, rather than because it is the most appropriate one to investigate a research problem (Pearce 2016: 47).

Lack of consistency in analytical techniques or protocols: Mixing results from studies by different labs, or conducted with different equipment or sample preparation, may make the results incomparable. Common problems are differences in detection limits and ranges of elements detected (Heginbotham et al. 2010; Pearce 2016).

Failure to report the details of method and analytical protocols: This is particularly prevalent in publications in conference proceedings or journals that do not specifically require such information (Pearce 2016: 49).

Dependence on "black-box" analysis: Particularly with the increasing use of portable X-ray Fluorescence (pXRF), many archaeologists fall prey to some manufacturers' claims that their proprietary software and built-in calibrations can provide accurate concentrations of major and trace elements with minimal intervention by the analyst, or none at all. These claims are illusory, especially for fairly complex alloys in which various peaks in the spectra can overlap, when there is secondary fluorescence, or when poor X-ray geometry, surface enrichment, or other factors may cause backscattering or differential absorption of X-rays.

Insufficient standards for calibration: Most of the methods that archaeometallurgists use to measure the compositions of artifacts yield spectra with peaks whose areas are related to the concentrations of different elements in the alloy (Fig. 13.4). To convert these areas into percentages or parts per million, it is necessary to calibrate the instrument by analyzing standards of known composition. However, since alloys with different proportions of elements have very complex relationships to the spectra produced, it is usually necessary to have a very large set of calibration standards with considerable ranges of contributions by the various elements expected in the archaeological metals.

Surface contamination, depletion, or enrichment: As museum curators and others tend, for good reason, to resist destructive analyses (see Chap. 9), many of the methods that archaeometallurgists favor, such as XRF, actually only analyze materials on or very slightly below the artifact's surface. This makes the results vulnerable to differences in methods used to clean the surface (or not), to contaminants that could originate from the burial environment or conservation practices, and to metallurgical practices (e.g., surface depletion or plating) that cause compositional differences between the surface and the main body of the artifact.

Other intra-artifact variations: As many analytical methods only analyze a small spot on the artifact's surface, or a small volume of removed material, it is possible that the results are not representative of the whole artifact (Pearce 2016: 47–48).

Effect of smelting and recycling on composition: Where the research goal is to identify ore sources, depletion of more volatile elements, such as arsenic, from repeated melting of the metal can have a large effect on its concentration in the surviving metal (Bray and Pollard 2012), so that it may be a poor match to the original ore or alloy.

Unavailability of reference data on ore sources or artifacts of known source: Sourcing artifacts, one of the major research themes in archaeometallurgy, depends heavily on being able to characterize the potential sources. Yet researchers have not always made these data sets available, while some journals are unwilling to publish such datasets (Pearce 2016: 49–50).

Some of the measures that analysts can take to protect the reliability and validity of their results include the following.

Selection of appropriate methods: Although availability and cost of analytical facilities are inevitable factors, analysts should select from available methods the ones that are most appropriate to the circumstances, considering a number of questions. Must they be non-destructive? Is surface enrichment likely? What elements or isotopes are necessary? What precision and detection limits are acceptable? Are fully quantitative results needed to answer the research questions, or only semi-quantitative ones?

Full description of analytical methods and protocols: As reproducibility from lab to lab is an important quality consideration, it is essential to ensure strict adherence to and full description of protocols that describe sample preparation, equipment, equipment settings, detection limits, calibration standards, and software procedures.

Attention to surface issues: It is helpful either to clean the portion of all surfaces that will be subject to analysis, or at least to conduct a study on a sample of the artifacts that compares uncleaned and cleaned surfaces to determine the extent of problems related to surface contamination, depletion or enrichment.

Use of standards: Analytical protocols should include calibrating equipment on a suite of standards that represent, as fully as possible, the ranges of elements or isotopes that occur in the artifacts. In some cases, it will be necessary to make custom standards, since commercially available ones do not necessarily have the right compositions.

Attention to inhomogeneity: Analyzing at least a sample of the artifacts in multiple locations helps to assess the extent to which inhomogeneity of artifacts might be a problem.

Interpretation of analytical spectra: It is essential not to abdicate responsibility for this to some automated system. Analysts must carefully "interrogate" the spectra to confirm the identification of peaks, resolve overlaps, and carry out the most appropriate calibrations.

Availability of raw data: Researchers should explore Open Access options for posting their raw data so that other researchers can make use of it in new research or to confirm previous results. Many universities host data repositories that are widely available online.

13.5 Summary

- Metal has unique characteristics as a raw material, relative to stone, ceramic and organic materials, including color, malleability, lustre, and relative ease of recycling
- The *chaînes opératoires* of metal artifacts are complex, and it is not straightforward to disentangle some aspects, as certain processes, including annealing, hammering or re-melting, can obscure or obliterate earlier processes
- As metal has usually been a valuable material, it sometimes provides special insights into prestige economies, commodity exchange, standards of value and market economy, where those existed
- Intercomparison of archaeometallurgical studies can be challenging as use of different methods, attention to different elements, isotopes or other attributes, restricted availability of source data, and variations in analytical quality or detection levels can greatly affect results

References Cited

Agarwal, B. K. (1991). *X-ray spectroscopy: An introduction* (2nd ed.). New York: Springer.

Aitken, M. J. (1974). *Physics and archaeology* (2nd ed.). New York: Interscience Publishers.

Artioli, G., Angelini, I., Kaufmann, G., Canovaro, C., Dal Sasso, G., & Villa, I. M. (2017). Long-distance connections in the copper age: New evidence from the alpine iceman's copper axe. *PLoS One, 12* (7), e0179263.

Bass, G. F. (1967). The ingots. Cape Gelidonya: A bronze age shipwreck. *Transactions of the American Philosophical Society, 57*(8), 52–83.

Bayley, J., & Rehren, T. (2007). Towards a functional and typological classification of crucibles. In S. La Niece, D. Hook, & P. Craddock (Eds.), *Metals and mines: Studies in archaeometallurgy* (pp. 46–55). London: Archetype Publications.

Bebber, M., & Eren, M. (2018). Toward a functional understanding of the North American old copper culture "technomic devolution". *Journal of Archaeological Science, 98*, 34–44.

Ben-Yosef, E. (2018a). Provenancing Egyptian metals: A methodological comment. *Journal of Archaeological Science, 96*, 208–215. https://doi.org/10.1016/j.jas.2018.06.001.

Ben-Yosef, E. (Ed.). (2018b). *Mining for ancient copper: Essays in memory of Professor Beno Rothenberg*. Winona Lake: Eisenbrauns.

Bevan, A. (2012). Spatial methods for analysing large-scale artefact inventories. *Antiquity, 86*(332), 492–506.

Bray, P. J., & Pollard, A. M. (2012). A new interpretative approach to the chemistry of copper-alloy objects: Source, recycling and technology. *Antiquity, 86*(333), 853–867.

Budd, P., & Taylor, T. (1995). The fairie smith meets the bronze industry: Magic *versus* science in the interpretation of prehistoric metal-making. *World Archaeology, 27*, 133–143.

Casey, P. J. (1986). *Understanding ancient coins: An introduction for archaeologists and historians*. London: Batsford.

Casey, P. J., & Reece, R. (Eds.). (1974). *Coins and the archaeologist*. Oxford: British Archaeological Reports.

Chase, W. T. (1991). *Ancient Chinese bronze art: Casting the precious sacral vessel*. New York: China Institute in America.

Chastain, M. L., Deymier-Black, A. C., Kelly, J. E., Brown, J. A., & Dunand, D. C. (2011). Metallurgical analysis of copper artifacts from Cahokia. *Journal of Archaeological Science, 38*(7), 1727–1736. https://doi.org/10.1016/j.jas.2011.03.004.

Childe, V. G. (1929). *The Danube in prehistory*. Oxford: Clarendon Press.

Childe, V. G. (1930). *The bronze age*. Cambridge: Cambridge University Press.

Collis, J. (1974). Data for dating. In J. Casey & R. Reece (Eds.), *Coins and the archaeologist* (pp. 173–183). Oxford: British Archaeological Reports.

Comelli, D., D'orazio, M., Folco, L., El-Halwagy, M., Frizzi, T., Alberti, R., et al. (2016). The meteoric origin of Tutankhamun's iron dagger blade. *Meteortics & Planetary Science, 51*(7), 1301–1309.

Cooper, H. K. (2011). The life (lives) and times of native copper in Northwest North America. *World Archaeology, 43*(2), 252–270. https://doi.org/10.1080/00438243.2011.581444.

Cooper, H. K. (2012). Innovation and prestige among northern hunter-gatherers: Late prehistoric native copper use in Alaska and Yukon. *American Antiquity, 77*(3), 565–590.

Cope, L. H. (1972). Surface-silvered ancient coins. In E. T. Hall & D. M. Metcalf (Eds.), *Methods of chemical and metallurgical investigation of ancient coinage* (pp. 261–278). London: Royal Numismatic Society.

Craddock, P. T. (1995). *Early metal mining and production*. Washington: Smithsonian Institution Press.

Craddock, P. T. (2001). From hearth to furnace: Evidences for the earliest metal smelting technologies in the eastern Mediterranean. *Paléorient, 26*, 151–165.

De Muynck, D., Cloquet, C., & Vanhaecke, F. (2008). Development of a new method for Pb isotopic analysis of archaeological artefacts using single-collector ICP-dynamic reaction cell-MS. *Journal of Analytical and Atomic Spectrometry, 23*(1), 62–71.

Deans, J. (1885). Copper money of the Haida. *The Museum, 1*(1), 9.

Doran, J. E., & Hodson, F. R. (1975). *Mathematics and computers in archaeology*. Cambridge, MA: Harvard University Press.

Ehrhardt, K. L. (2009). Copper working technologies, context or use, and social complexity in the eastern woodlands of native North America. *Journal of World Prehistory, 22*, 213–235.

Esty, W. W. (1986). Estimation of the size of a coinage: A survey and comparison of methods. *Numismatic Chronicle, 146*, 185–215.

Fairbank, W. (1962). Piece-mould craftsmanship and Shang bronze design. *Oriental Art, 16*, 9–15.

Franklin, U. M. (1999). *The real world of technology* (2nd ed.). Toronto: Anansi.

Gale, N. H. (1979). Lead isotopes and archaic Greek silver coins. *Archaeo-physika, 10*, 194–208.

Gale, N. H., Stos-Gale, Z. A., Maliotis, G., & Annetts, N. (1997). Lead isotope data from the Isotrace Laboratory, Oxford: *Archaeometry* data base 4, ores from Cyprus. *Archaeometry, 39*(1), 237–246.

Gentner, W., Müller, O., Wagner, G. A., & Gale, N. H. (1978). Silver sources of Archaic Greek coinage. *Die Naturwissenschaften, 65*(6), 273–284.

Goffer, Z. (2007). *Archaeological chemistry* (2nd ed.). New York: Wiley.

Gopher, A., & Tsuk, T. (1991). *Ancient gold: Rare finds from the Nahal Qanah Cave*. Jerusalem: Israel Museum.

Greenfield, H. (2013). 'The fall of the house of flint': A zooarchaeological perspective on the decline of chipped stone tools for butchering animals in the Bronze and Iron Ages of the southern Levant. *Lithic Technology, 38*(3), 161–178.

Halsey, J. (1996). Without forge or crucible: Aboriginal native American use of metals and metallic ores in the Eastern Woodlands. *The Michigan Archaeologist, 14*(1), 1–58.

Heeb, J. (2014). *Copper shaft-hole axes and early metallurgy in South-Eastern Europe*. Oxford: Archaeopress.

Heginbotham, A., Bezur, A., Bouchard, M., Davis, J., Eremin, K., Frantz, J., et al. (2010). An evaluation of inter-laboratory reproducibility for quantitative XRF of historic copper alloys. In P. Markikian, C. Chemello, C. Watters, & P. Hull (Eds.), *Metal 2010, proceedings of the interim meeting of the ICOM-CC metal working group, Charleston, South Carolina, USA* (pp. 244–254). Clemson: Clemson University.

Hodder, I., & Orton, C. (1976). *Spatial analysis in archaeology*. Cambridge: Cambridge University Press.

Hodder, I., & Reece, R. (1977). A model for the distribution of coins in the western Roman Empire. *Journal of Archaeological Science, 4*, 1–18.

Hodges, H. (1964). *Artifacts: An introduction to early materials and technology*. London: J. Baker.

Hörz, G., & Kallfass, M. (1998). Pre-Columbian metalworking in Peru—Ornamental and ceremonial objects from the royal tombs of Sipan. *JOM, 50*(12), 8–16.

Hosler, D. (1995). Sound, color and meaning in the metallurgy of ancient West Mexico. *World Archaeology, 27*(1), 100–115.

Hosler, D., & Macfarlane, A. (1996). Copper sources, metal production, and metals trade in late Postclassic Mesoamerica. *Science, 273* (5283), 1819–1824.

Hulme, E. W. (1937). Prehistoric and primitive iron smelting, part I: From meteoric iron to 1000 B.C. *Transactions of the Newcomen Society, 18*, 181–192.

Jopling, C. F. (1989). The coppers of the Norwest Coast Indians: Their origin, development, and possible antecedents. *Transactions of the American Philosophical Society, 79*(1), 1–164.

Kemmers, F., & Myrberg, N. (2011). Rethinking numismatics. The archaeology of coins. *Archaeological Dialogues, 18*(1), 87–108.

Kenoyer, M. (2012). Measuring the Harappan world: Insights into the indus order and cosmology. In I. Morley & C. Renfrew (Eds.), *The archaeology of measurement: Comprehending heaven, earth and time in ancient societies* (pp. 106–122). Cambridge: Cambridge University Press.

Killick, D. J. (2015). Archaeometallurgy as archaeology. A key-note lecture. In A. Hauptmann & D. Modaressi-Tehrani (Eds.), *Archaeometallurgy in Europe III: Proceedings of the 3rd international conference* (pp. 295–300). Bochum: Deutsches Bergbau-Museum.

Killick, D. J., & Fenn, T. (2012). Archaeometallurgy: The study of preindustrial mining and metallurgy. *Annual Review of Anthropology, 41*, 559–575.

Kimes, T., Haselgrove, T., & Hodder, I. (1982). A method for the identification of regional cultural boundaries. *Journal of Anthropological Archaeology, 1*, 113–131.

La Niece, S., Hook, D., & Craddock, P. (Eds.). (2007). *Metals and mines: Studies in archaeometallurgy*. London: Archetype Publications.

Laing, L. R. (1969). *Coins and archaeology*. London: Weidenfeld and Nicolson.

Lo Schiavo, F., Muhly, J. D., Maddin, R., Giumlia-Mair, A. (Eds.) (2009). Oxhide ingots in the Central Mediterranean. Rome: CNR — Instituto di Studi sulle Cività dell'Egeo e del Vicino Oriente.

Lockyear, K. (2012). Dating coins, dating with coins. *Oxford Journal of Archaeology, 31*(2), 191–211.

Loehr, M. (1968). *Ritual vessels of bronze age China*. New York: The Asia Society.

Märcher, M. (2012). Six early modern Danish coin hoards. *Numismatic Chronicle, 172*, 323–332.

Martin, S. R. (1990). *Wonderful power: The story of ancient copper working in the Lake Superior basin*. Detroit: Wayne State University Press.

McGovern, W. E. (1980). Missing die probabilities, expected die production and the index figure. *American Numismatic Society Museum Notes, 25*, 209–223.

Moorey, P. R. S. (1988). The chalcolithic hoard from Nahal Mishmar, Israel, in context. *World Archaeology, 20*(2), 171–189.

Muhly, J. D., Wheeler, T. S., & Maddin, R. (1977). The Cape Gelidonya shipwreck and the bronze age metals trade in the eastern Mediterranean. *Journal of Field Archaeology, 4*(3), 353–362.

Ottaway, B. (1994). *Prähistorische archäometallurgie*. Espelkamp: Marie Leidorf.

Ottaway, B. (2002). Towards interpretive archaeometallurgy. In M. Bertelheim, E. Pernicka, & R. Krause (Eds.), *Die anfänge der metallurgie in der alten welt* (pp. 7–13). Rahden: Marie Leidorf.

Ottaway, B., & Strahm, C. (1975). Swiss Neolithic copper beads: Currency, ornament or prestige items? *World Archaeology, 6*(3), 307–321.

Pearce, M. (2016). Archaeology and archaeometallurgy: Some unresolved areas in the interpretation of analytical data. *Science and Technology of Archaeological Research, 2*(1), 46–53.

Pole, L. M. (2010). The hammers of Mawu: Ironworking traditions in the Togo Hills, Ghana. *African Archaeological Review, 27*(1), 43–78.

Pollard, A. M., & Bray, P. (2014). Chemical and isotopic studies of ancient metals. In B. W. Roberts & C. P. Thornton (Eds.), *Archaeometallurgy in global perspective: Methods and syntheses* (pp. 217–238). New York: Springer.

Pollard, A. M., Bray, P. J., & Gosden, C. (2014). Is there something missing in scientific provenance studies of prehistoric artefacts? *Antiquity, 88*(340), 625–631.

Price, T. D., & Burton, J. H. (2012). *An introduction to archaeological chemistry*. New York: Springer.

Rahmstorf, L. (2012). The concept of weighting during the bronze age in the Aegean, the Near East and Europe. In I. Morley & C. Renfrew (Eds.), *The archaeology of measurement: Comprehending heaven, earth and time in ancient societies* (pp. 88–105). Cambridge: Cambridge University Press.

Roberts, B. W., & Thornton, C. P. (Eds.). (2014). *Archaeometallurgy in global perspective: Methods and syntheses*. New York: Springer.

Rothenberg, B. (Ed.). (1978). *Archaeometallurgy: Chalcolithic copper smelting*. London: Institute for Archaeo-Metallurgical Studies.

Sapiro, D., & Webler, B. (2016). Fabrication of a bronze age sword using ancient techniques. *JOM, 68*(12), 3180–3185.

Schmidt, P. R. (1996). Reconfiguring the Barongo: Reproductive symbolism and reproduction among a work association of iron smelters. In P. R. Schmidt (Ed.), *The culture and technology of African iron production* (pp. 74–127). Gainesville: University Press of Florida.

Shearman, F. (1990). Metals. In S. Bradley (Ed.), *A guide to the storage, exhibition and handling of antiquities, ethnographia and pictorial art* (British Museum Occasional Paper 66) (pp. 18–26). London: The British Museum.

Shugar, A. N., & Mass, J. L. (Eds.). (2012). *Handheld XRF for art and archaeology*. Leuven: Leuven University Press.

Smith, C. S. (1975). Metallurgy as a human experience. *Metallurgical Transactions A, 6A*(3), 603–623.

Starr, C. G. (1970). *Athenian coinage, 480–449 B.C*. Oxford: Clarendon Press.

Stos-Gale, Z. A., Gale, N. H., & Annetts, N. (1996). Lead isotope data from the Isotrace laboratory, Oxford: *Archaeometry* data base 3, ores from the Aegean, part 1. *Archaeometry, 38*(2), 381–390.

Tadmor, M., Kedem, D., Begemann, F., Hauptmann, A., Pernicka, E., & Schmitt-Strecker, S. (1995). The Nahal Mishmar hoard from the Judean Desert: Technology, composition, and provenance. *Atiqot, 27*, 95–148.

Timberlake, S. (2007). The use of experimental archaeology/archaeometallurgy for the understanding and reconstruction of early bronze age mining and smelting technologies. In S. La Niece, D. Hook, & P. Craddock (Eds.), *Metals and mines: Studies in archaeometallurgy* (pp. 27–36). London: Archetype Publications.

Tucci, A., Sayavongkhamdy, T., Chang, N., & Souksavatdy, V. (2014). Ancient copper mining in Laos: Heterarchies, incipient states or post-state anarchists? *Journal of Anthropology and Archaeology, 2* (2), 1–15.

Tushingham, A. D., Franklin, U. M., & Toogood, C. (1979). *Studies in ancient peruvian metalworking: An investigation of objects from the Museo Oro del Peru Exhibited in Canada in 1976–77 under the title, 'Gold for the Gods'*. Toronto: Royal Ontario Museum.

Tylecote, R. F. (1992). *A history of metallurgy* (2nd ed.). London: The Institute of Metals.

Villa, I. M. (2009). Lead isotopic measurements in archaeological objects. *Archaeological and Anthropological Sciences, 1*, 149–153. https://doi.org/10.1007/s12520-009-0012-5.

Wadsworth, J. (2015). Archaeometallurgy related to swords. *Materials Characterization, 99*, 1–7.

Wayman, M. L. (1985). Native copper: Humanity's introduction to metallurgy? In M. L. Wayman (Ed.), *All that glitters: Readings in historical metallurgy* (pp. 3–6). Montreal: Canadian Institute of Mining and Metallurgy.

Woolf, A. D., Law, T., Yu, H. Y. E., Woolf, N., & Kellogg, M. (2010). Lead poisoning from use of bronze drinking vessels during the late Chinese Shang dynasty: An *in vitro* experiment. *Clinical Toxicology, 48*, 757–761.

Zhanwei, Y. (2008). The Shang bronze foundry-site at Xiaomintun in Anyang city. *Chinese Archaeology, 8*(1). https://doi.org/10.1515/CHAR.2008.8.1.16.

Zimmer, G. F. (1917). The use of meteoric iron by primitive man. *Nature, 98*, 350–352.

14 Bone and Shell Tools

> *The pointed horn of the deer furnishes the ready-made dagger, lance-head, and harpoon; the incisor tooth of the larger rodents supplies a more delicately edged chisel than primitive art could devise; and the very process of fracturing the bones of the larger mammalia in order to obtain the prized marrow, produces the splinters and pointed fragments which an easy manipulation converts into bodkins, hair-pins, and needles*
>
> Wilson (1876: 96–97)

Although the antiquity of their first use is still a matter of debate, bone and shell have long served as materials for a wide array of tools (Krzyszkowska 1990; Langley 2016; St.-Pierre and Walker 2007). Bone has been particularly important for use in piercing technologies, while shell has been a common material for those involving cutting, indenting, and scraping (Szabó et al. 2014). Both have served widely for use in personal ornaments, especially beads and pendants, and fasteners for clothing. Bone needles were a notable requisite for making that clothing before metal ones became available. Antler has often been important for soft hammers for lithic production, and the tines make good pressure-flakers. In addition, human and animal body parts, especially bones, were sometimes curated and modified for symbolic uses and ritual (e.g., Hargrove et al. 2015).

14.1 The *Chaînes Opératoires* of Bone and Shell Technology

Bone and shell technologies often take advantage of the existing characteristics of unmodified bone and shell. For example, bone awls and projectile points are often made from long bones whose natural shape is fairly close to the shape of the desired end-product, or employ a condyle or articulation as a butt or handle (Fig. 14.1), while cylindrical shells such as *Dentalium* are easy to segment into beads, and other kinds of shell have shapes that are often conducive to an intended artifact design. In fact, some tools require little or no modification at all; soft hammers for stone-tool manufacture, for example, might show little or no modification apart from battering from use, while unmodified clam shells might only show use wear from use as a scoop or potting tool.

14.1.1 Bone, Shell and Ivory as Raw Materials

Bone consists of both organic and inorganic components. The inorganic materials that cells in bones deposit include calcium phosphate ($Ca_3(PO_4)_2$), calcium carbonate ($CaCO_3$), magnesium, fluorine, chlorine and iron.

Shelled molluscs generate an exoskeleton consisting mainly of calcium carbonate (mainly calcite, aragonite, or both), along with very small amounts of protein and sometimes an outer proteinaceous layer (periostracum). Shell's hardness is similar to that of limestone, but with even greater tensile strength (Currey 1976, 1979; Currey and Taylor 1974), making it highly suitable for a wide range of cutting and scraping tools. There are many different ways that aragonite or calcite is deposited in shell, leading to a variety of microstructures that differ in their tensile strength and type of crack propagation (Szabó 2008: 129–130; Wilbur and Saleuddin 1983: 257).

Ivory, the material that makes up elephant and mammoth tusks, consists mainly of dentine, a calcified epithelial tissue that is also one of the major components of teeth (Krzyszkowska 1990). It is a porous, bone-like and somewhat yellow material whose major component is the mineral, hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$), along with calcium phosphate, collagen and other materials.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_14

Fig. 14.1 Bone awls created by reducing the distal ends of sheep or goat metapodials to a point but leaving the proximal articulation as a sort of handle (S. Rhodes)

14.1.2 Preparation of the Material

Most archaeologists who have attempted to replicate bone tools have selected fairly fresh bone with most of the flesh already removed, or used stone flakes to scrape off remaining flesh and sinew, and sometimes periosteum, and then have soaked the bone in water for periods of 1 day to 5 weeks. Soaking softens it and makes it easier to work (Wojtczak and Kerdy 2018: 799). In one study, preparation of bone blanks took 4 hours after one day of soaking, but only 1.5 hours after 3–5 weeks of soaking (Arrighi et al. 2016: 151).

14.1.3 Shaping the Material

Much as with lithics, bone and shell technologies are reductive, meaning that shaping tools involves removal of material. In fact, this analogy has led to characterization of bones used as tool material as "cores" and pieces removed from them as "blanks" (Betts 2007; Morrison 1986; Smith and Poggenpoel 1988; Zhang et al. 2018). However, while percussive flaking can be among the reductive techniques used to shape bone and thick shell, incision, sawing, drilling, and abrasion are much more common for both bone and shell.

Percussion: Somewhat as with lithic tools, striking bones or thicker shells sharply with a hammerstone serves to break them into smaller pieces, including sharp slivers that might make good blanks for pointed tools. Breaking bones sometimes involved a version of bipolar or hammer-and-anvil percussion, but more careful percussion can also be used to crush edges of bone blanks to help shape them. Striking thick shells more carefully with a hammerstone can also serve to make holes, especially in combination with a chisel (Bar-Yosef Mayer 1997: 98), and it is even possible to pressure-flake the edges of some kinds of shell (Szabó 2008: 132).

Wedge-splitting, or groove-and-splinter: It is possible to split bones longitudinally by first incising or cutting a groove, and then driving a chisel-like wedge into the bone with a hammer (Fig. 14.2). As the wedge widens the opening, cracks will tend to propagate along the groove, thus helping to control the split (Morrison 1986: 110).

Snapping: Since many bones are elongated, snapping is a useful method for reducing them or their fragments into shorter lengths, sometimes by applying sharp downward pressure at the two ends while the middle of a long bone is supported on a fulcrum, such as a stick or stone. Similar methods can be used to snap larger shells into pieces.

Cutting or sawing: Sawing with a stone or metal edge serves to reduce the shafts of long bones or tubular shells into shorter sections, although causing considerable wear to saw edges. It is also possible to saw them with bone or wooden tools, or even string, with quartz sand as an abrasive (Bar-Yosef Mayer 1997: 98). Sawing can also be used to make notches.

Incision or scoring: Much as with sawing, incising V-shaped grooves at desired locations can be a first step to reducing the shaft of a long bone into sections. The toolmaker could then finish the separation of those sections by snapping, with the underside of the grooved section on a fulcrum, and the incisions guiding the fracture propagation. However, incision can also be used longitudinally on cortical bone to ease splintering and isolate pieces that can be used for needles or points, and also to make decorative patterns or tally marks.

Scraping: Scraping a bone blank with a stone edge is useful for removing material quickly as well as for fine-tuning the shape or tapering it to a point.

Abrasion: Rubbing against a grinding stone is useful for sharpening bone points, removing imperfections in snapped or sawn sections, or thinning needles. Where it is necessary to remove a lot of bone, however, it is better to precede grinding with scraping (Arrighi et al. 2016: 151). Grinding is also useful for wearing down a convex surface of shells to form a hole or flatten it for attachment to clothing. Naturally abraded shells collected on sand beaches can already show holes that make them suitable as beads or pendants, but the abrasion may be evenly distributed over their surfaces, rather than focused on one area, as seen on intentionally-ground shells (Bar-Yosef Mayer 1997: 97–98). It may also be possible to distinguish naturally abraded from artifactual shell or shells selected for use as beads by comparing the extent of abrasion on modern and "fossil" examples with that on shells from archaeological assemblages (d'Errico et al. 2009; Figs. 14.3 and 14.4).

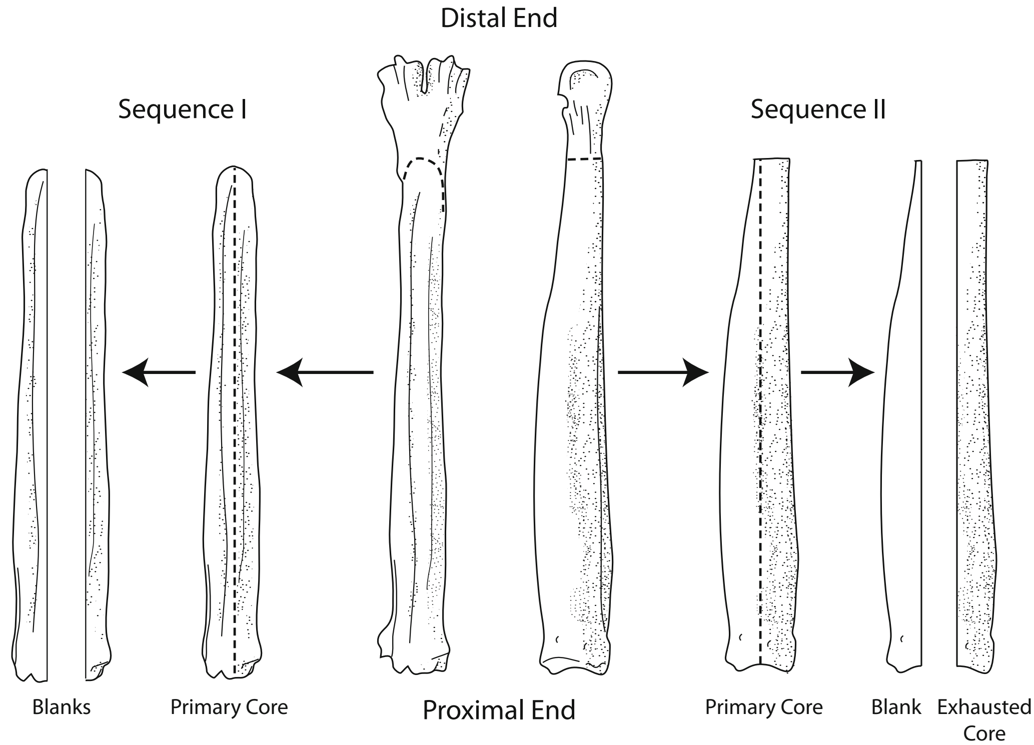


Fig. 14.2 An example of primary reduction based on longitudinal splitting, in this case, of a caribou metatarsal, first in the coronal plane, posterior of the distal condyle, followed by splitting the posterior portion into medial and lateral halves (left) or (right) along sagittal plane. (After Morrison 1986: 110)

Fig. 14.3 Ordinal scales for scoring the extent of wear and perforation on the dorsal and ventral sides of Nassarius shells. (After d'Errico et al. 2009: 16054)

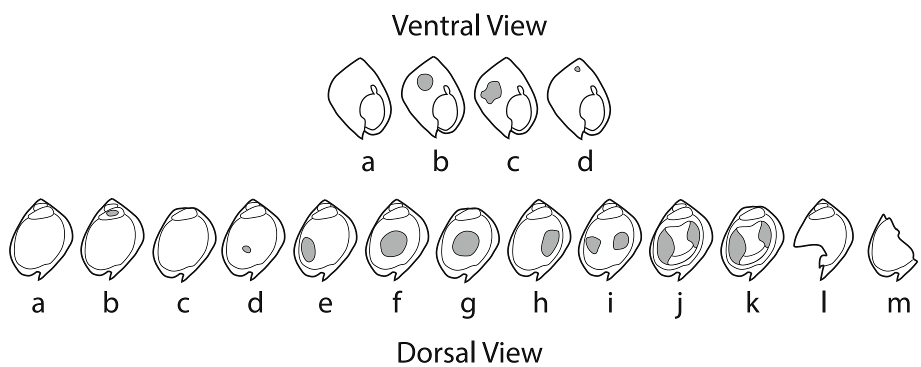
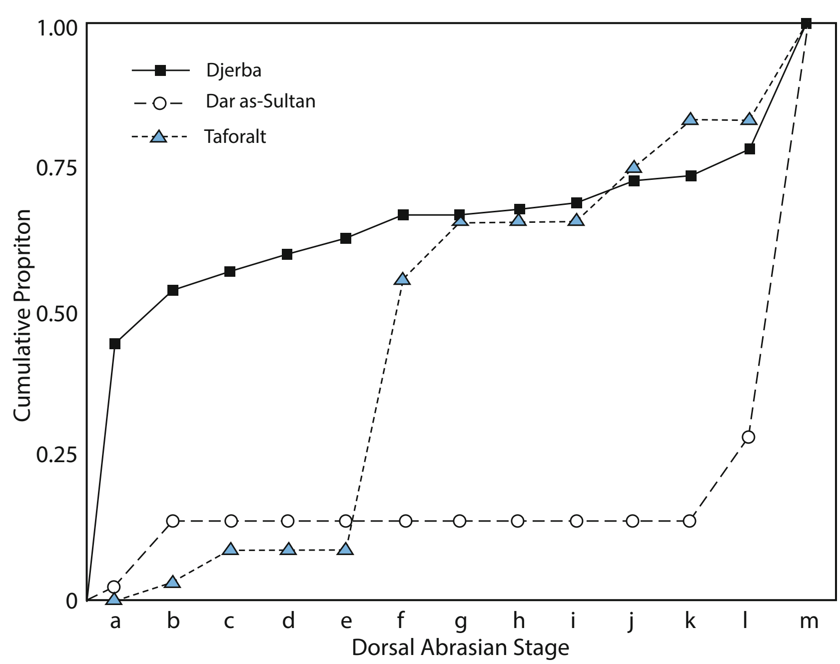


Fig. 14.4 Cumulative frequency of dorsal abrasion and perforation in a sample of modern, beach-abraded Nassarius shells from Djerba, Tunisia, from a fossil assemblage at Dar as-Sultan, Morocco, and an archaeological assemblage from Taforalt, Morocco (data from d'Errico et al. 2009: 16054), using the ordinal scale in Fig. 14.3. Note how the archaeological sample from Taforalt shows a high proportion of shells at stage f, while the fossil assemblage is mostly stage m



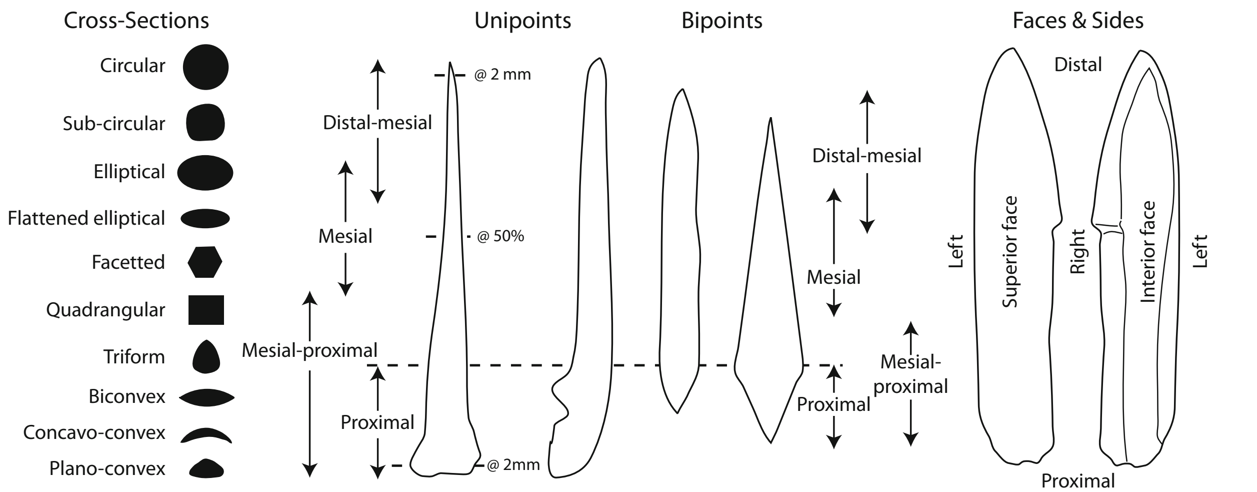


Fig. 14.5 An example of a classification of cross-sectional shape (left) and segmentation rules for bone points. (After Langley 2018: 8)

Drilling: Biconical (i.e., from both sides) drilling with a stone drill is useful for piercing bone tools to allow suspension of pendants or bull-roarers, to sew buttons or toggles onto clothing, or to create eyes on needles or fishhooks. It was also common to pierce harpoon heads to allow the attachment of lines or lanyards (Cristiani and Boric 2016).

14.2 Types and Anatomies of Bone and Shell Tools

Bone, shell and ivory have been used for such a wide array of tools that it is impossible to summarize them thoroughly in a single chapter. What follows highlights just a few broad classes of artifacts and technologies that take advantage of these materials' unique properties.

14.2.1 Bone Awls, Needles and Points

Bone has frequently been a material of choice for making various narrow, pointed tools. In such tools, much as with lithic points, it is conventional to orient them with the most pointed end upward and described as the distal end (Fig. 14.1). The other end or base is the proximal end, which may also be pointed or rounded but sometimes retains much or all of the condyle or epiphysis of a long bone. Note that, even though we call this the proximal end of the tool, it is not necessarily the proximal end of a bone. Archaeologists often classify bone points and other pointed tools with attributes that include the anatomical part used, how much of that part remains, and how the part was shaped (e.g., Figs. 14.5 and 14.6). More specialized points, such as harpoon heads, require more complex classifications and segmentation rules that take barbs and potential piercings into account (Fig. 14.7). For tools with a somewhat flattened cross-section (flattened elliptical, biconvex, or concavo-convex), one may distinguish a dorsal or superior face from a ventral or inferior face. The fairly flat profiles of ribs and some other bones make them very useful for making shuttles, shed sticks and combs for weaving, while cortical bone has long been a material of choice for sewing needles (Legrand 2007, 2008; Lyman 2015). Common measures for all such tools include maximum length, axial length (from distal point to center of the base at the proximal end), maximum width, and widths at the middle of the long axis (medial width), halfway from the medial line to the distal end (distal-medial width), and halfway from the medial line to the proximal end (proximal-medial width).

14.2.2 Fishhooks

Prehistoric fishhooks were often made from bone or shell. The anatomy of fishhooks includes a notched or pierced head for attachment of the line, the shank, the bow (or bend), the point, and often a barb (Fig. 14.8). To make ratio-scale measurements of these features, it is necessary to establish

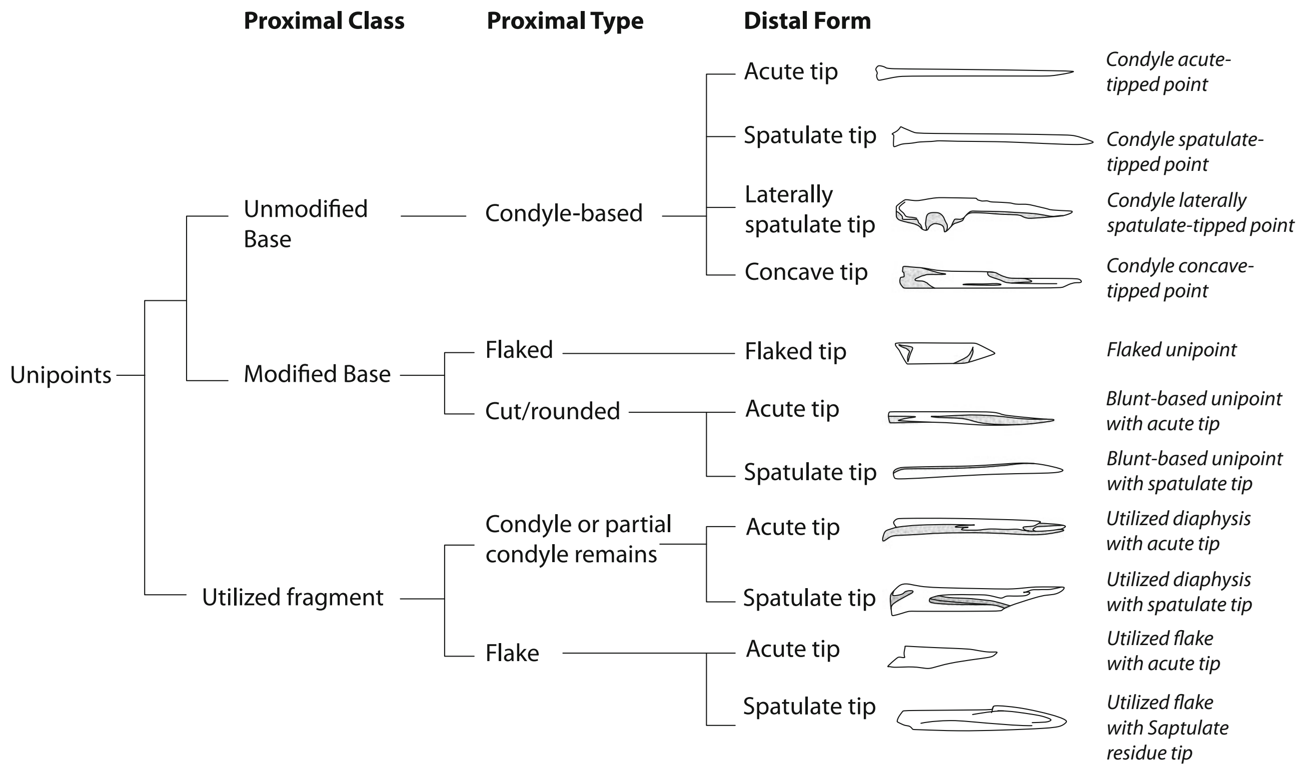


Fig. 14.6 Classification of Australian bone points represented as a taxonomy. (Y. Salama, modified from Langley 2018: 9)

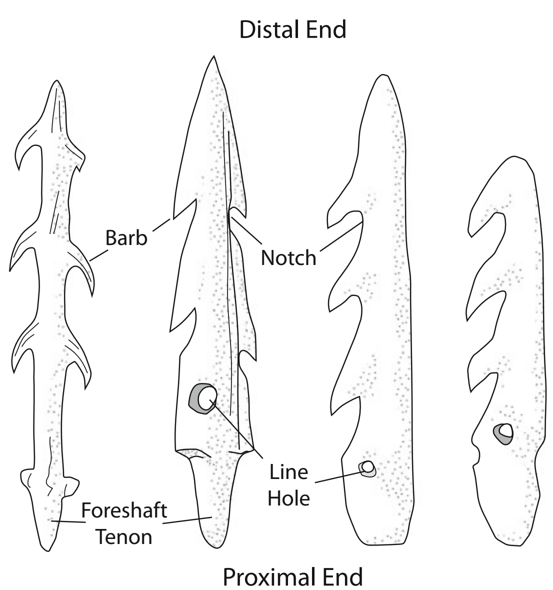


Fig. 14.7 Harpoon points and some of their terminology

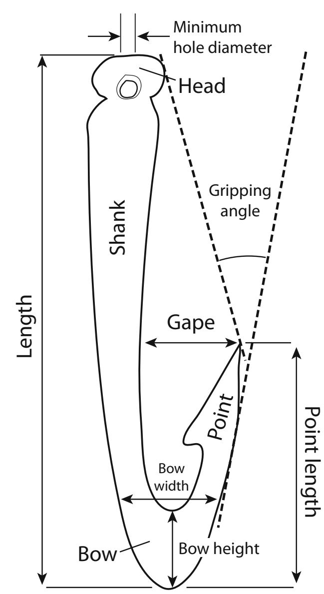


Fig. 14.8 Segmentation and standard measurements for bone fishhooks. (Modified after Olson et al. 2008: 2816)

the hook's axis, a line that extends from the bottom of the bow and perpendicularly bisects a line from the tip of the point to the dorsal edge of the shank. The length of the hook should be measured parallel to this axis. Other possible measures include the maximum shank width and thickness, height of the bow, maximum width of the bow, height of the point, width of the barb, the gape, and the gripping angle.

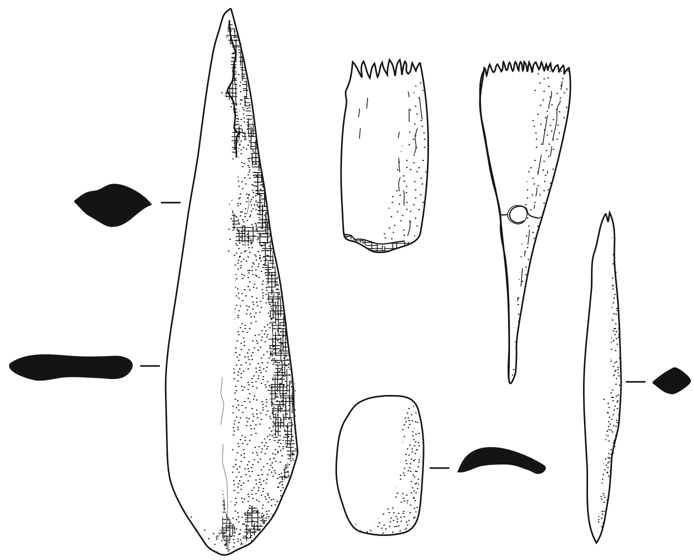


Fig. 14.9 A variety of bone tools including "ribs" for shaping, thinning, combing and indenting pottery. (After Gibson and Woods 1997: 46–47)

14.2.3 Potting Tools

Bone and shell are highly suitable for use as the tools that potters use for turning, scraping, impressing, incising and rouletting pottery (Fig. 14.9; see pp. 192–193). With fairly minimal alteration by snapping, cutting and grinding, bones with relatively flat cross-sections, such as ribs, make excellent trimming knives and scrapers; it is no coincidence that potters call some of these tools "ribs." Bones with notched edges, and unmodified or minimally modified bivalve shells also make excellent serrated scrapers or combs for cases where the potter wants to create a textured surface, and they are also useful for impressing patterns on vessel surfaces (see p. 202).

14.2.4 Beads and Pendants

Both bone and shell have been common materials for personal ornaments since the Palaeolithic (Álvarez Fernández and Jöris 2007; Balme and Morse 2006). Shell, in particular, has been very important for the production of ornament-like prestige goods, sometimes called "primitive money," which often circulated far from their place of origin, as in the Mississippian of the American Southeast, or the Kula Ring of the Trobriand Islands (Brain and Phillips 1996; Malinowski 1922; Swadling and Bence 2016; Trubitt 2003, 2005; Yang 2019). These range from items that had recognizable monetary properties to ones that had symbolic value and were used in exchanges but did not have all the properties of Western money (Dalton 1965; Einzig 1966; Szabó 2018). Certain cowries (especially *Monetaria moneta*) served as a form of currency in many parts of the Old World as late as the colonial period, and were sometimes replicated in other materials, even precious metals (Golani 2014; Yang 2011, 2019).

It is fairly easy to make some kinds of beads by cutting sections from birds' long bones or cylindrical shells, such as *Dentalium*, or by grinding down portions of rounder shells. Other kinds of shells frequently served as pendants, which only required piercing them for attachment, unless the maker wanted to alter the overall shape. However, some kinds of beads, such as the small wampum beads prevalent in eastern North America prior to the colonial import of glass beads, required considerable labor to manufacture (Ceci 1982: 100; Wilcox 1976).

Pierced artifacts of bone or shell have frequently been useful as buttons and toggles for the closure of clothing, as well as for decorative (and sometimes symbolically important) elements on clothing. Disk-shaped pieces of shell have been particularly useful for buttons, while sawn or ground-down cowrie (*Cypraeidae* spp.) shells have often served as decorative or symbolic elements sewn onto clothing.

14.2.5 Bull-Roarers, Whistles and Flutes

Many bones are potentially useful for making aerophones (instruments that make sounds by vibrating air).

Relatively dense bone from diaphyses of long bones could have been used for making bull-roarers. These look like tapered pendants but are actually the aerodynamic part of a musical instrument that makes a whistling or roaring sound when someone swings the "pendant" rapidly on a long thong (Armstrong 1936; Harding 1974; Morley 2013).

Among the earliest candidates for whistles are pierced phalanges. Phalanges of deer and other animals with piercings on one side would produce a high-pitched sound if someone blew across the hole, much like blowing across a bottle's opening. Archaeologists have identified examples of possible phalangeal whistles from as early as the European Aurignacian, ca. 40,000–26,000 BC (Megaw 1960; Morley 2013). However, the most parsimonious explanation for many of the earliest reputed whistles is that carnivore gnawing or some other natural process caused the piercings (Chase 1990; d'Errico et al. 2003). An alternative hypothesis for the culturally modified ones is that they are figurines, perhaps representing women or babies (Caldwell 2009).

Any hollow long bone that is open at both ends will also produce sounds if someone blows across one end, at an angle to the edge, but generally archaeologists do not recognize bones as flutes unless they are pierced with at least one hole. The purpose of the hole or holes is to allow a musician to alter the wavelength of the sound by placing fingers over some of the holes. This changes the length of the column of vibrating air inside the tube (or air spring). While a simple tube only produces one tone, a tube with holes can thus produce multiple tones that are higher than the tone with all holes closed (Fletcher and Rossing 1998; Nederveen 1998). It is possible that people played monotonal flutes long before they started perforating them to make more tones, but it would be difficult to demonstrate that archaeologically.

As with the phalangeal whistles, it is debatable whether all, or even any, of the earliest pierced bone shafts are flutes, rather than bones with accidental piercings (d'Errico et al. 2003; Diedrich 2015). For example, an alleged flute of probable Middle Palaeolithic age from Haua Fteah in Libya has one hole and part of a possible second hole that are consistent with carnivore gnawing (Davidson 1991). Similarly, d'Errico et al. (1998) have argued that the alleged 60,000-year-old Middle Palaeolithic flute from Divje Babe Cave in Slovenia has holes that are identical in their size, distribution and shape to ones in cave-bear accumulations where there was no evidence of human occupation.

While these early examples now seem doubtful, much clearer examples of bone flutes occur in Upper Palaeolithic contexts in Europe. The site of Isturitz, France, for example, has more than 20 convincing examples of bird-bone flutes, some with as many as four holes, from contexts 35,000–20,000 years old (Fig. 14.10; Buisson 1990; García Benito et al. 2016; Lawson and d'Errico 2002).

Distinguishing flutes and whistles from bones holed by natural processes depends on examination of a number of attributes. Cut-marks may indicate intentional adjustment of the shaft's length, while cross-sectional shape of the holes is a very important clue (conical or biconical holes indicate drilling, while cylindrical, depressed, and sharp holes can result from puncture by teeth), and some kinds of holes result from gastric acid in a carnivore's digestive tract. The character and distribution of use-wear traces can also be instructive; polish and striations can result from placement of hands and fingers, especially near the holes. Pitting all over surfaces and tooth indentations or holes opposite the major holes is indicative of animal gnawing (Davidson 1991; d'Errico and Villa 1997).

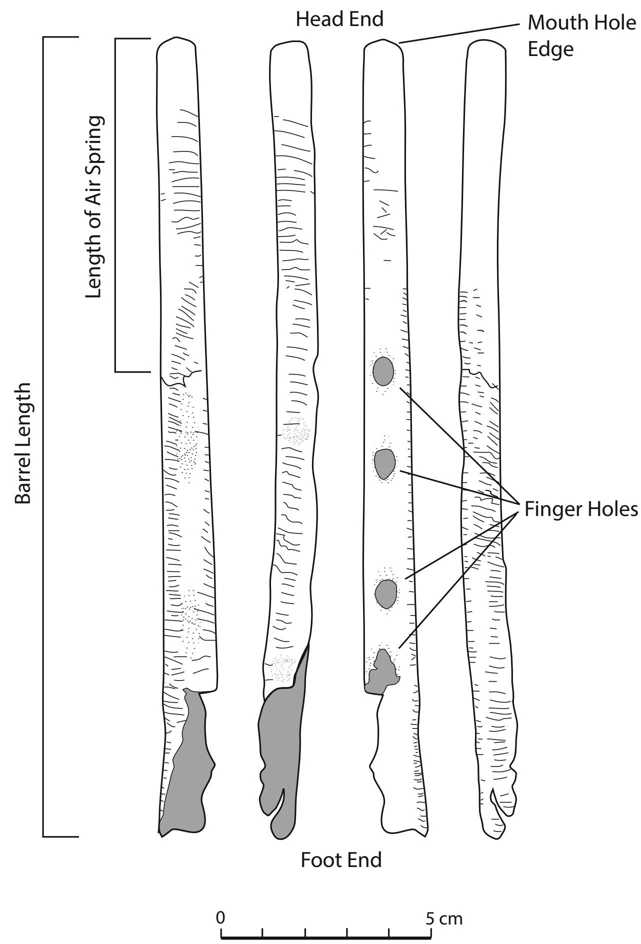


Fig. 14.10 A simple, open-tube flute, illustrated with an example from Isturitz (modified from Buisson 1990). Ignoring harmonics, the wavelength of a standing wave in the bore (interior diameter) would be approximately double the distance between the mouth and the first open finger hole (length of air spring)

14.2.6 Bone, Ivory, and Shell as Decorative Inlay

In addition to their use as substantial tool components, these materials have often served for decorative inlay in other materials, such as wood. When wooden boxes and cabinets inlaid with them have decayed away, they may be the only traces of such artifacts.

For example, inlay has often been a feature of luxury containers and furniture, making it a good candidate for an indirect measure of wealth or status. In a Late Bronze Age context at the site of Megiddo in Israel, for example, excavators found more than 200 incised and carved ivory plaques with Phoenician designs influenced by Egyptian art, including images of griffons, sphinxes, and lotus blossoms. Most of these appear to have ornamented furniture, but some appear to be from boxes and toilet articles, while some are combs and one is a gaming board (Kantor 1956; Loud 1939; Wilson 1938). Some of the plaques have piercings that show that at least some were attached with pins or nails.

Case Study

Ivory for the Sea Woman

Robert McGhee (1977) draws attention to certain classes of artifacts of the Thule culture of the Alaskan and Canadian Arctic that tend to be made preferentially from marine-mammal teeth (colloquially described as "ivory"), marine-mammal bone, or caribou antler, and wonders if functional constraints and availability were the only factors affecting selection of these materials. Could symbolic factors also have been important?

In the collections McGhee presents, all of the un-barbed and single-barbed arrowheads, a type that recent Inuit used to hunt caribou, are made from antler (table 14.1). With the exception of the Lady Franklin Point site, harpoon heads were mostly made from ivory or sea-mammal bone, as were many of the components of harpoons and their associated gear.

McGhee summarizes by saying that ivory appears to have been associated with hunting sea mammals and birds, life on the sea ice, and women and, along with sea-mammal bone, is negatively associated with caribou hunting. He finds that purely functional considerations do not satisfactorily account for these patterns. Ivory could just as easily have been worked to make caribou arrowheads as harpoon heads, and the fact that a few harpoon foreshafts were made from antler demonstrates that antler was not unsuited for that purpose, just rarely chosen. He draws on structuralism to propose sets of symbolic oppositions, land:sea, summer:winter, man:woman, and antler:ivory or "land is to sea as summer is to winter, as man is to woman and as antler is to ivory."

With reference to Table 14.1, how strongly do you think the data support McGhee's hypothesis? What statistical test might you use to evaluate his hypothesis?

Table 14.1 Counts of selected artifact classes made from walrus ivory or sea-mammal teeth, sea-mammal bone, and antler, pooled from Thule sites Walakpa, Lady Franklin Point, Nunguvik, Cumberland Sound, and Silumiut (data in McGhee 1977; means rounded to significant digits). McGhee suggests that ivory and sea-mammal bone tend to be associated with equipment that historic Inuit used to hunt birds, including multi-barbed arrowheads, dart prongs, and bolas. Arrowheads that men used for hunting caribou are routinely made of antler. Finally, he also notes that certain artifact classes that were mainly associated with women among nineteenth-century Inuit, such as sewing equipment, combs, and pendants, are of ivory

Artifact category	Ivory	Bone	Antler	Row means
Caribou arrowheads	0	0	528	176
Harpoon heads	56	62	106	75
Bird arrowheads	2	5	8	5
Bird dart sideprongs	7	1	6	5
Foreshaft sockets	3	13	5	7
Harpoon foreshafts	13	15	3	10
Bolas balls	28	32	0	20
Bird-woman figures	16	0	0	5.3
Pendants	14	0	0	4.7
Combs	8	0	0	3
Needle cases	3	0	0	1
Thimble holders	2	0	0	1
Column means	13	11	55	

14.3 Use Wear on Bone and Shell Tools

While not as common as studies of lithic or even ceramic use wear, research on use wear on bone and shell tools also provides important evidence for activities involving these tools (Stemp et al. 2016). Because bone and shell have physical characteristics that are quite different from those of stone, use traces can be quite different as well. For shell tools, Cuenca-Solana et al. (2017) conclude that use traces even vary by the species of shell. Some studies have successfully distinguished traces on bone and antler caused by uses for projectiles, hide-working, scaling fish, piercing, pottery forming, weaving, sewing, net-making, and sewing reeds (e.g., Arrighi et al. 2016; Buc 2011; Legrand 2007; LeMoine 1997; Shipman and Rose 1988; Soffer 2004; Wojtczak and Kerdy 2018).

As with use wear on lithics, most studies depend on expertise and qualitative assessment of wear, striations and polish, but some researchers have begun to use quantitative textural analysis to characterize use wear on bone tools and increase the reliability and validity of results (e.g., Martisius et al. 2018).

14.4 Summary

- Bone, antler, tooth, ivory, and shell have properties that make them highly suitable for manufacture of a wide variety of tools and ornaments.
- Many of such tools take advantage of the natural shapes of the material. For example, ribs are already elongated and somewhat flat, and long bones have cylindrical shafts, while bivalves are concave and have sharp and sometimes serrated edges.
- The *chaînes opératoires* of such tools have some things in common with those of stone tools, in that they involve reductive technology and sometimes percussion flaking.
- Technological parallels with production of groundstone tools are particularly notable, with such techniques as incising, sawing, grinding and polishing being important.

- Bone has been particularly important for the production of pointed tools, such as needles, projectile points, fishhooks, and awls. Consequently, some of the same terminology as used for pointed lithics is applicable.
- Many anatomical parts, notably teeth and small shells, are highly suitable as preforms for ornaments, such as beads.

References Cited

Álvarez Fernández, E., & Jöris, O. (2007). Personal ornaments in the early Upper Palaeolithic of western Eurasia. An evaluation of the record. *Eurasian Prehistory, 5*, 31–44.

Armstrong, A. L. (1936). A bull-roarer of Le Moustier age from Pin Hole Cave, Creswell Cags, Derbyshire. *The Antiquaries Journal, 16*(3), 322–323.

Arrighi, S., Bazzanella, M., Boschin, F., & Wierer, U. (2016). How tomake and use a bone 'spatula.' An experimental program based on the Mesolithic osseous assemblage of Galgenbühel/Dos de la Forca (Salurn/Salorno, BZ, Italy). *Quaternary International, 423*, 143–165.

Balme, J., & Morse, K. H. (2006). Shell beads and social behaviour in Pleistocene Australia. *Antiquity, 80*, 799–811.

Bar-Yosef Mayer, D. E. (1997). Neolithic shell bead production in Sinai. *Journal of Archaeological Science, 24*, 97–111.

Betts, M. W. (2007). The Mackenzie Inuit whale bone industry: Raw material, tool manufacture, scheduling, and trade. *Arctic, 60*(2), 129–144.

Brain, J. P., & Phillips, P. (1996). *Shell gorgets: Styles of the late prehistoric and protohistoric southeast*. Cambridge: Peabody Museum of Archaeology and Ethnology, Harvard University.

Buc, N. (2011). Experimental series and use-wear in bone tools. *Journal of Archaeological Science, 38*(3), 546–557.

Buisson, D. (1990). Les flutes paléolithiques d'Isturitz (Pyrénées-Atlantiques). *Bulletin de la Société Préhistorique Française, 87*(10/12), 420–433.

Caldwell, D. (2009). Palaeolithic whistles or figurines? A preliminary survey of pre-historic phalangeal figurines. *Rock Art Research, 26*(1), 65–82.

Ceci, L. (1982). The value of wampum among the New York Iroquois: A case study in artifact analysis. *Journal of Anthropological Research, 38*, 97–107.

Chase, P. G. (1990). Sifflet du Paléolithique moyen (?). Les implications d'un coprolithe de coyote actuel. *Bulletin de la Société Préhistorique Française, 87*(6), 165–167.

Cristiani, E., & Boric, D. (2016). Mesolithic harpoons from Odmut, Montegnegro: Chronological, contextual, and techno-functional analyses. *Quaternary International, 423*, 166–192.

Cuenca-Solana, D., Gutiérrez-Zugasti, I., & González-Morales, M. R. (2017). Use-wear analysis: An optimal methodology for the study of shell tools. *Quaternary International, 427*, 192–200.

Currey, J. D. (1976). Further studies on the mechanical properties of mollusc shell material. *Journal of Zoology, 180*, 445–453.

Currey, J. D. (1979). The effect of drying on the strength of mollusc shells. *Journal of Zoology, 188*, 301–308.

Currey, J. D., & Taylor, J. D. (1974). The mechanical behavior of some molluscan hard tissues. *Journal of Zoology, 173*, 395–406.

d'Errico, F., & Villa, P. (1997). Holes and grooves: The contribution of microscopy and taphonomy to the problem of art origins. *Journal of Human Evolution, 33*, 1–31.

d'Errico, F., Villa, P., Pinto Llona, A. C., & Ruiz Idarraga, R. (1998). A Middle Palaeolithic origin of music? Using cave-bear bone accumulations to assess the Divje Babe I bone 'flute.'. *Antiquity, 72*(275), 65–79.

d'Errico, F., Henshilwood, C., Lawson, G., Vanhaeren, M., Tillier, A. M., Soressi, M., et al. (2003). Archaeological evidence for the emergence of language, symbolism and music: An alternative multidisciplinary perspective. *Journal of World Prehistory, 17*(1), 1–70.

d'Errico, F., Vanhaeren, M., Barton, N., Bouzouggar, A., Mienis, H., Richter, D., et al. (2009). Additional evidence on the use of personal ornaments in the Middle Paleolithic of North Africa. *Proceedings of the National Academy of Science, 106*, 16051–16056.

Dalton, G. (1965). Primitive money. *American Anthropologist, 67*(1), 44–65.

Davidson, I. (1991). The archaeology of language origins: A review. *Antiquity, 65*, 39–48.

Diedrich, C. G. (2015). 'Neanderthal bone flutes': Simply products of Ice Age spotted hyena scavenging activities on cave bear cubs in European cave bear dens. *Royal Society Open Science, 2*(4). https://doi.org/10.1098/rsos.140022.

Einzig, P. (1966). *Primitive money in its ethnological, historical and economic aspects* (2nd ed.). London: Pergamon Press.

Fletcher, N. H., & Rossing, T. D. (1998). *The physics of musical instruments* (2nd ed.). New York: Springer.

García Benito, C., Alcolea, M., & Mazo, C. (2016). Experimental study of the aerophone of Isturitz: Manufacture, use-wear analysis and acoustic tests. *Quaternary International, 421*, 239–254.

Gibson, A., & Woods, A. (1997). *Prehistoric pottery for the archaeologist*. London: Leicester University Press.

Golani, A. (2014). Cowrie shells and their imitations as ornamental amulets in Egypt and the Near East. *Polish archaeology in the Mediterranean, 23*(2), 71–83.

Harding, J. R. (1974). The bull-roarer in history and in antiquity. *African Music, 5*(3), 40–42.

Hargrove, E. A., Shermer, S. J., Hedman, K. M., & Lillie, R. M. (Eds.). (2015). *Transforming the dead: Culturally modified bone in the prehistoric Midwest*. Tuscaloosa: The University of Alabama Press.

Kantor, H. (1956). Syro-Palestinian ivories. *Journal of Near Eastern Studies, 15*, 153–174.

Krzyszkowska, O. (1990). *Ivory and Related Materials: An Illustrated Guide* (Classical Handbook 3). London: Institute of Classical Studies.

Langley, M. C. (Ed.). (2016). *Osseous projectile weaponry: Towards an understanding of Pleistocene cultural variability*. Dordrecht: Springer.

Langley, M. C. (2018). Establishing a typology for Australian pointed bone implements. *Australian Archaeology, 84*(2), 164–180. https://doi.org/10.1080/03122417.2018.1509541.

Lawson, G., & d'Errico, F. (2002). Microscopic, experimental and theoretical re-assessment of Upper Paleolithic bird-bone pipes from Isturitz, France: Ergonomics of design, systems of notation and the origins of musical tradition. In E. Hickmann, A. Kilmer, & R. Eichmann (Eds.), *Archäologie früher klangerzeugung und tonordnung: Vorträge des 2. Symposiums der internationalen studiengruppe musikarchäologie im kloster michaelstein* (pp. 17–23). Rahden: M. Leidorf.

Legrand, A. (2007). *Fabrication et utilization de l'Outillage en Matières Osseuses du Néolithique de Chypre: Khirokitia et Cap Andreas-Kastros*. Oxford: British Archaeological Reports.

Legrand, A. (2008). Neolithic bone needles and vegetal fibres working: Experimentation and use-wear analysis. In N. Skakun & L. Longo (Eds.), *Prehistoric technology 40 years later: Functional studies and the russian legacy* (pp. 445–450). Oxford: British Archaeological Reports.

LeMoine, G. (1997). *Use wear on bone and antler tools from the Mackenzie Inuit*. Oxford: British Archaeological Reports.

Loud, G. (1939). *The Megiddo Ivories*. Chicago: The Oriental Institute and University of Chicago Press.

Lyman, R. L. (2015). North American Paleoindian eyed bone needles: Morphometrics, sewing, and site structure. *American Antiquity, 80* (1), 146–160.
Malinowski, B. (1922). *Argonauts of the Western Pacific*. London: Routledge and Kegan Paul.
Martisius, N. L., Sidéra, I., Grote, M. N., Steele, T. E., McPheron, S. P., & Schulz-Kornas, E. (2018). Time wears on: Assessing how bone wears using 3D surface texture analysis. *PLoS ONE, 13*(11), e0206078. https://doi.org/10.1371/journal.pone.0206078.
McGhee, R. (1977). Ivory for the sea woman: The symbolic attributes of a prehistoric technology. *Canadian Journal of Archaeology, 1*, 141–149.
Megaw, J. V. S. (1960). Penny whistles and prehistory. *Antiquity, 34*, 6–13.
Morley, I. (2013). *The prehistory of music: Human evolution, archaeology, and the origins of musicality*. Oxford: Oxford University Press.
Morrison, D. (1986). Inuit and Kutchin bone and antler industries in northwestern Canada. *Canadian Journal of Archaeology, 10*, 107–125.
Nederveen, C. J. (1998). *Acoustical aspects of wind instruments*. DeKalb: Northern Illinois University.
Olson, C., Limburg, K., & Söderblom, M. (2008). Stone Age fishhooks—how were they dimensioned? Morphology, strength test, and breakage pattern of Neolithic bone fishhooks from Ajvide, Gotland, Sweden. *Journal of Archaeological Science, 35*, 2813–2823.
Shipman, P., & Rose, J. J. (1988). Bone tools: An experimental approach. In S. L. Olsen (Ed.), *Scanning electron microscopy in archaeology* (pp. 303–335). Oxford: British Archaeological Reports.
Smith, A. B., & Poggenpoel, C. (1988). The Technology of Bone Tool Fabrication in the South-Western Cape, South Africa. *World Archaeology, 20*(1), 103–115.
Soffer, O. (2004). Recovering perishable technologies through use wear on tools: Preliminary evidence for Upper Paleolithic weaving and net making. *Current Anthropology, 45*, 407–413.
St.- Pierre, C. G., & Walker, R. B. (Eds.). (2007). *Bones as tools: Current methods and interpretations in worked bone studies*. Oxford: Archaeopress.
Stemp, W. J., Watson, A., & Evans, A. A. (2016). Topical review: Surface analysis of stone and bone tools. *Surface topography: Metrology and properties, 4*(1). http://iopscience.iop.org/article/10.1088/2051-672X/4/1/013001.
Swadling, P., & Bence, P. (2016). Changes in kula valuables and related supply linkages between the Massim and the South Papuan coast between 1855 and 1915. *Archaeology in Oceania, 51*(supp. 1), 50–60.
Szabó, K. (2008). Shell as a raw material: Mechanical properties and working techniques in the tropical Indo-West Pacific. *Archaeofauna, 17*, 125–138.
Szabó, K. (2018). Shell money and context in Western Island Melanesia. In L. Carreau, A. Clark, A. Jelinek, E. Lilje, & N. Thomas (Eds.), *Pacific presences: Oceanic art and European museums* (Vol. 2, p. 2538). Leiden: Sidestone Press.
Szabó, K., Dupont, C., Dimitrijevic, V., Gómez Gastélum, L., & Serrand, N. (Eds.). (2014). *Archaeomalacology: Shells in the archaeological record*. Oxford: Archaeopress.
Trubitt, M. B. D. (2003). The production and exchange of marine shell prestige goods. *Journal of Archaeological Research, 11*(3), 243–277.
Trubitt, M. B. D. (2005). Crafting marine shell prestige goods at Cahokia. *North American Plains Archaeologist, 26*, 249–266.
Wilbur, K. M., & Saleuddin, A. S. M. (1983). Shell formation. In A. S. M. Saleuddin & K. M. Wilbur (Eds.), *Physiology (Part 1). The mollusca* (Vol. 4, pp. 235–287). New York: Academic Press.
Wilcox, U. (1976). The manufacture and use of wampum in the Northeast: Description and manufacture. *The Bead Journal, 3*(1), 10–19.
Wilson, S. D. (1876). *Prehistoric man: Researches into the origin of civilisation in the Old and New World* (Vol. 1). London: Macmillan & Co.
Wilson, J. A. (1938). The Megiddo ivories. *American Journal of Archaeology, 42*(3), 333–336.
Wojtczak, D., & Kerdy, M. (2018). They left traces. Preliminary analyses of micro-wear traces on bone and antler tools from Sutz-Lattrigen Aussen, Lake Bienne, Switzerland. *Journal of Archaeological Science: Reports, 17*, 798–808.
Yang, B. (2011). The rise and fall of cowrie shells: The Asian story. *Journal of World History, 22*(1), 1–25.
Yang, B. (2019). *Cowrie shells and Cowrie money: A global history*. New York: Routledge.
Zhang, S., Doyon, L., Zhang, Y., Xing, G., Chen, F., Guan, Y., et al. (2018). Innovation in bone technology and artefact types in the late Upper Palaeolithic of China: Insights from Shuidonggou Locality 12. *Journal of Archaeological Science, 93*, 82–93.

15 Archaeological Animal Remains

The obvious advantage of research with fauna is that . . . their form is not a product of any mental templates or designs for living. Therefore it becomes possible to set forth a completely culture-free taxonomy of bones. Any variability observed in the relative frequencies of anatomical parts among archaeological sites must derive from the dynamics of their use . . . partitioning, segmenting, and differentially distributing the segments. of animal anatomy

Binford (1978: 11)

Zooarchaeology is the study of the bones, teeth, shells, and other traces of animals from archaeological contexts as evidence for the lives of past people. Archaeologists can use these sources of evidence to identify and understand ancient environments, hunting and herding behaviors, dietary preferences, seasonality of site occupations, use of animals for non-food purposes, such as fur or traction, and even social behaviors and the mobility of people or livestock across territories (Rackham 1994; Russell 2012; Steele 2015). Zooarchaeology also has subdisciplines, such as archaeomalacology (the study of archaeological molluscs, Bar-Yosef Mayer 2005), and archaeoentomology (the study of archaeological insect remains, Buckland 2000; Buckland et al. 2004; Panagiotakopulu 2000).

This chapter will introduce general aspects of zooarchaeology, including the classification of animals and the main body parts of some animal classes, the processes that affect their representation in archaeological deposits, and some additional ways to count and measure aspects of faunal assemblages besides taxonomic abundance and diversity (see Chap. 7). It will go on to illustrate with examples some of the common research problems that archaeologists try to investigate through animal remains. Although humans are animals too, and some of the things covered in this chapter would also apply to human remains, the chapter will not explicitly deal with human remains or mortuary archaeology.

As usual, readers should consult more comprehensive sources for the topics briefly reviewed here (e.g., Albarella and Trentacoste 2011; Giovas and LeFebvre 2018; Lyman 2008; O'Connor 2008).

15.1 Types of Faunal Remains

Bones are not the only faunal remains that archaeologists are likely to find. Bones are often the best preserved and most obvious evidence of animals with bony skeletons, but it is sometimes possible to recover fragments of cartilage, skin and hair, antler, hoof, horn, muscle tissue, stomach contents, and coprolites (fossil feces), especially in waterlogged or very dry sites. Teeth are often preserved even in sediments that are not conducive to bone preservation, and even dental calculus can reveal ancient bacteria and food traces. Fragments of birds' eggs (Collins and Steele 2017; Keepax 1981) and even parts of insect exoskeletons and parasite eggs (e.g., Panagiotakopulu 2000) sometimes survive for observant archaeologists to find them. The shell of land snails, carapaces and claws of freshwater and marine crabs, sponge spicules, and marine shells often occur in archaeological contexts. An indirect but potentially important source of evidence can be nests or distinctive traces of insect or worm tunnelling. Archaeologists have even identified places where livestock were fenced in by distinctive spherulites — tiny calcitic spheres formed in the digestive tracts of some animals — in their dung. Tiny but very important evidence comes from otoliths (ear stones) of fish. There is also very important evidence from chemical residues on artifacts, stable isotopes in bone and teeth, and increasingly preserved collagen and ancient DNA (aDNA). This chapter, however, will place emphasis on bones, teeth, otoliths and mollusc shell.

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47992-3_15) contains supplementary material, which is available to authorized users.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_15

15.2 Taphonomy and Site-Formation Processes

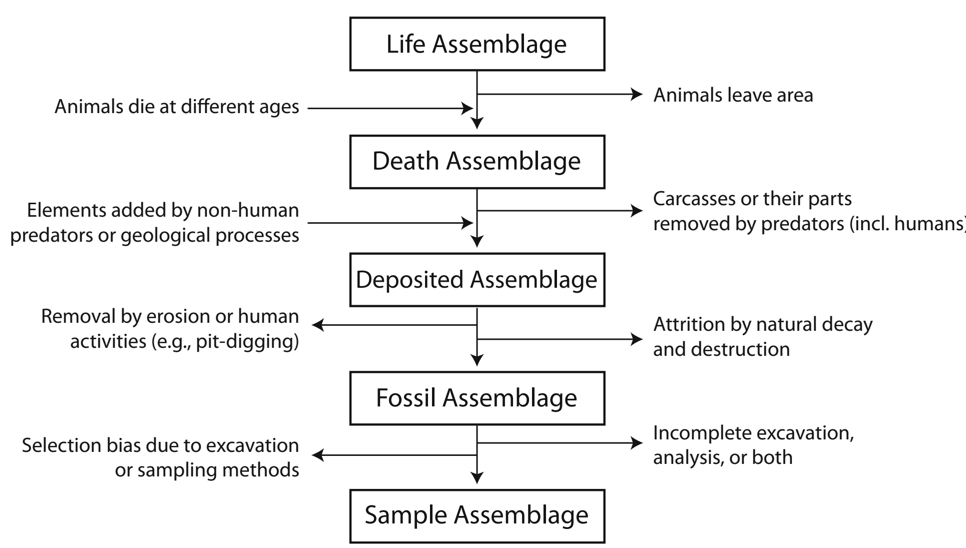
Assuming that faunal remains that we find represent ancient populations of animals or food presents major challenges to validity. Like other archaeological remains, faunal remains are sensitive to the processes that led to their eventual discovery by archaeologists (Kendall et al. 2018). The sets of remains that archaeologists find are not at all equivalent to the sets of animals, or even the sets of bones or shells, that once existed on a site or its environs. We do not directly study populations of animals, but rather *samples* of bones, bone fragments, and other kinds of fossils that exist in the present (see Chap. 6), at the end of a sequence of processes that includes the death and only sometimes the preservation and discovery of some parts of some of those animals. There are so many processes that affect different kinds of animals and different body parts in different ways that our modern sample is only a distorted remnant of that original population. Zooarchaeologists find it useful to view this as a sequence of selections that gradually reduced a large original population of living creatures to a non-random sample of their remains (Fig. 15.1).

Taphonomy is the study of the processes by which living animals in the biosphere eventually become (through fossilization) part of the earth's rock or lithosphere (Efremov 1940; Gifford 1981; Lyman 1994; 2014; Orton 2012). Taphonomists try to identify the processes that affected the composition of a faunal assemblage in the hope of reconstructing the structure of the population of animals at the time they died, or to understand the factors that caused humans to select certain animals or animal parts for food or some other purpose. Taphonomy has much in common with what archaeologists call site-formation processes (Schiffer 1976, 1987), and includes various kinds of theoretical, geoarchaeological, analogical, actualistic, and experimental work, such as ethnoarchaeological observation (Binford 1981) and experiments with bone trampling or carnivore consumption (Collins 2010; Gifford 1981; Moclán et al. 2019; Payne and Munson 1985). Taphonomic processes are the actions whose forces modify the physical characteristics and distribution of animal carcasses and tissues, while a taphonomic effect is the trace of such a process (Lyman 1994: 35).

Typically, zooarchaeologists distinguish among five types of populations or assemblages (Holtzman 1979; Klein and Cruz-Uribe 1984; Clark and Kietzke 1967; Meadow 1980):

- The **Life Assemblage**: This is the living community or population of animals from which the remains are derived.
- The **Death Assemblage**: This is the population of carcasses that results when members of the life assemblage die. Except when there is a catastrophic kill of the whole community, we would not expect the proportions of animals of various ages, sexes, and health status in the assemblage to be the same as in the life assemblage for the simple reason that death is not equally probable for animals in these categories. For example, the death assemblage is likely to include disproportionately high numbers of very young and old individuals. This may be the population whose size (N) the Peterson index (pp. 115–116) estimates.
- The **Deposited Assemblage**: This is the population of carcasses or their body parts that were deposited on an archaeological site through the actions of humans, non-human predators, scavengers, or such agents as water flow, wind and gravity. Notably, some of these actions, including hunting practices, choice of body parts to take back to camp, and disposal practices, are likely to deposit different kinds of animals or body parts in different places on the landscape, while parts of a single carcass could end up in two or more deposited assemblages.
- The **Fossil Assemblage**: This is a subset of the deposited assemblage that consists of those animal parts that survive in the site's deposit until their potential discovery by an archaeologist. Environmental conditions, such as the acidity of the surrounding sediment, site erosion, and damage

Fig. 15.1 Model of the taphonomic sequence beginning with a living population of animals at a moment or short interval of time and ending with an archaeological sample, along with some of the processes that can affect the probability of animal parts' survival from one stage to the next (arrows pointing outward) and a few processes that could add to or affect assemblage content (inward arrows). Generally, the assemblages are smaller the lower they are in this sequence



by trampling or scavengers have a substantial effect on the character of this assemblage. In terms of sampling theory (Chap. 6), this (or some spatial volume that contains it) is the population that archaeologists actually sample.

- The **Sample Assemblage**: This is the portion of the fossil assemblage that someone excavated or collected and then analyzed. In fact, often the analyzed sample is only a subset of what was collected (a subsample). Needless to say, archaeologists' field and laboratory methods, such as how carefully they excavated and whether and how they screened for small items, have a huge impact on the character and size of this sample, whether it is a probability sample or not.

At each stage from the life assemblage to the sample assemblage, the absolute number of potentially observable bones or other animal parts decreases. Unfortunately for us, different kinds of remains experience taphonomic effects at different rates. As a result, the distribution of remains of different kinds changes so that they no longer reflect their ratios at earlier stages in the sequence (but see "ratio-of-ratios," p. 119). As with archaeobotanical remains, archaeologists have spent much effort trying to understand or compensate for the effects of differential preservation during these transitions, especially those between the deposited, fossil, and sample assemblages. Failure to do so results in uncontrolled errors that would likely jeopardize our interpretation of the evidence.

Case Study

Density-mediated Attrition of Bone

Of the processes implicated in the taphonomic sequence just outlined, zooarchaeologists have paid particular attention to the effect and sources of fragmentation (see also Chap. 7), which can occur at multiple stages following the animal's death. For example, mammal bones can be broken during butchering and marrow extraction, or mollusc shells broken to extract the mollusc, even before the deposited assemblage, while post-depositional processes, archaeological activities and suboptimal storage can fragment them still further.

Zooarchaeologists have noticed that certain kinds of faunal remains are more prone to fragmentation than others, and one of the significant controlling factors appears to be bone density (Lyman 2013).

Lyman (1984) notes that a simple correlation between bone density and apparent survivorship of various skeletal elements in a number of previously studied faunal assemblages is suggestive but does not prove causality, given questions about validity. He found inconsistencies among correlations in different assemblages that should be similar, indicating that we should not put much confidence in them, no matter the statistical significance of individual cases. In archaeological samples, we would also expect human behaviors, not just density, to be responsible for some differences in bone survivorship. Furthermore, previous studies defined bone "density" variously, although most often as mass divided by volume, and used different methods for measuring it, varying especially in how they accounted for voids in the bone if they measured volume by displacement in water.

Lyman instead used a photon densitometer, a type of scanner that measures the attrition of a photon beam as it passes through a material to yield an indirect measure of mineral mass per unit of length (linear density, g/cm) rather than bulk density (g/cm^3). The scanner can also measure bone width, thus allowing measurement of areal density (g/cm^2).[1] The densitometer does not measure bulk density, but it is possible to derive estimates of bulk density from the other measures. In this early study, Lyman used densitometer measurements at carefully-defined scan sites on skeletal elements of deer, domestic sheep, and a pronghorn antelope to test the statistical hypothesis that the number of identified specimens (NISP, see Chap. 7) is a function of density:

$$NISP = f(D_i)$$

Where D_i is the mean density of the ith skeletal element of a deer, sheep or antelope, and he examined this relationship for all three kinds of density in fauna from several different assemblages.

Lyman found that estimated bulk density was correlated with survivorship in four of seven paleontological assemblages and three of 12 archaeological or ethnographic assemblages. Six of the cultural and one of the paleontological assemblages showed correlations between survivorship and linear density, while one paleontological and four cultural assemblages showed correlations with areal density. Seven assemblages showed no correlation between survivorship and any of the density measures.

Lyman explains the paleontological correlations with bulk density by the assemblages' origin in carnivore dens, where hyenas or wolves would have had leisure to gnaw the bones and destroy the less dense ones. The failure of some of the cultural assemblages to show correlation to bulk density, but sometimes to show correlations with linear or areal density, may be due to cultural selection of anatomical parts for their food value.

Lyman suggested that future research would improve on this study by taking such factors as the size, shape, and porosity of bones into account, a suggestion that did come to fruition (e.g., Faith et al. 2007; Lam et al. 1998; Lyman 2013; Stiner 2002).

[1] Lyman uses idiosyncratic terminology for two of these measures, "bone mineral" for g/cm and "linear density" for g/cm^2.

15.3 Animal Taxonomy and Anatomy

While Binford's (1978: 11) assertion that zoological taxonomy is culture-free is dubious, a European taxonomy has become standardized for use in scientific discourse. Carl von Linné, an eighteenth-century Swedish anatomist, developed the basic structure of the taxonomic classification of organisms that biologists still use today. Its hierarchy begins with domain at the first level, with successive subdivisions at the levels of kingdom, phylum, class, order, family, genus, and species (Fig. 15.2). In the kingdom Animalia, Mammalia is a class of the phylum Chordata (animals that have a spinal chord). Because there is an evolutionary, that is, historical, basis for the hierarchy in this taxonomy (even though von Linné himself was ignorant of this evolutionary history), the taxonomic relationships somewhat track relationships of descent. Animals descended from the same ancestor tend to share quite a few characteristics. This is very helpful to zooarchaeologists trying to identify animals on the basis of fragmentary evidence. When they do not have enough macroscopic evidence to identify a fossil to species, often they can at least identify it to a particular family or genus because of these familial similarities. Bone fragments that are too small for macroscopic identification can sometimes be identified through microscopic features of tissue structure (e.g., Sawada et al. 2014) or, more expensively, by ancient DNA. The category to which we assign a fossil is its taxon (plural, taxa), whether that is a species, genus, or some more general category.

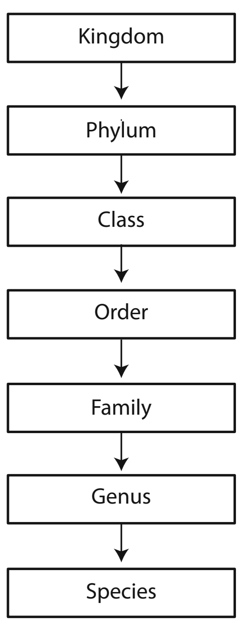


Fig. 15.2 Linnaean taxonomic system for Animalia

15.3.1 Mammalian Anatomy

Through their evolutionary history, mammals (members of the class, Mammalia) have much the same skeletal parts, even though their skeletons vary in the total number of bones and, of course, the shapes of those bones. For describing parts of skeletons, zooarchaeologists employ the same segmentation rules and analytical elements that zoologists use (see Fig. 15.3). In addition to the names for whole bones and teeth, there are also terms for their parts and surfaces: epiphysis (bone end) and diaphysis (bone shaft or centrum), proximal and distal ends (toward or away from the body), medial and lateral position (near or away from the midline of the body), anterior and posterior (fore- and rear parts of the body), dorsal (toward the back) and ventral (toward the stomach) for bones, and buccal (cheek-side) and lingual (tongue side) for teeth. There are also names for groups of bones within the structure of the skeleton. These include the axial (skull, vertebral column, ribs, and sternum) and appendicular (fore-limbs or arms, hind-limbs, pectoral girdle or shoulder region, and pelvic girdle or hip region) parts.

Because zooarchaeologists are often interested in how past people butchered and consumed animals, they sometimes supplement these basic segmentation rules with ones that butchers would use. For example, recent butchers in England would subdivide a pig's carcass into such portions as loin, belly, shank, forehock, and gammon (Davis 1987).

Bone and cartilage are the supporting tissues in vertebrate organisms (animals with a backbone). Bone is a living tissue that consists of cells and their products, blood vessels, and nerves. The inorganic materials that the cells deposit include calcium phosphate ($Ca_3(PO_4)_2$), calcium carbonate ($CaCO_3$), magnesium, fluorine, chlorine and iron. These are the materials that sometimes survive in fossil assemblages long after the organism is dead and most of the organic components have decayed away. In addition, the fossilization process itself deposits minerals from the surrounding sediment while gradually converting the bone into a stone, as in paleontological fossils. Bone cells also deposit fibrous protein material, similar to that found in cartilage, of which collagen is very important. All red and many white blood cells are produced in the long bones. Cartilage is durable and generally more flexible than bone. It consists of cells in a matrix of protein, carbohydrates, and fibres.

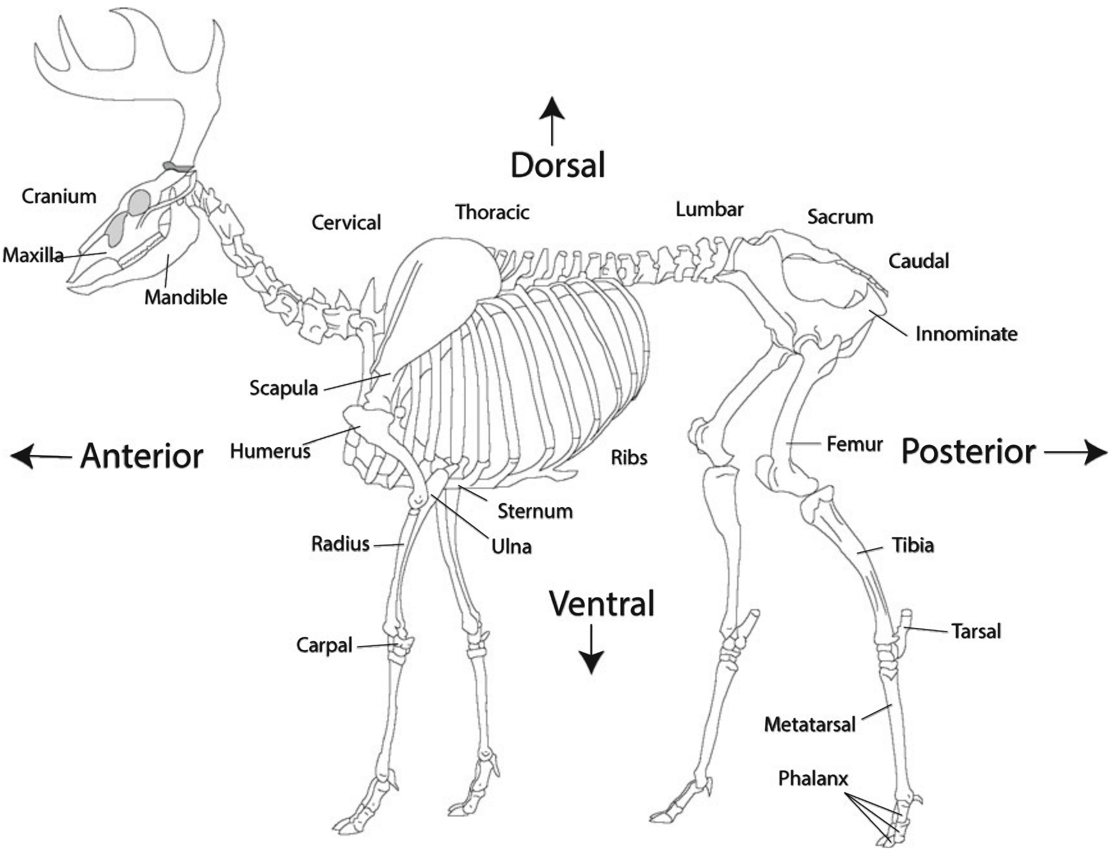


Fig. 15.3 Left lateral view of the skeleton of a cervid (deer) skeleton to illustrate the major skeletal parts and orientations in ungulates (W. Wadsworth)

Bone is formed by ossification over the lifetime of a vertebrate animal. Early in the animals's life, the "bones" are really composed mainly of cartilage. Ossification begins midway between the proximal and distal ends of a bone and later occurs at the two epiphyses (Fig. 15.4). The bone-forming cells remain alive within the matrix materials that they secrete, and the first bone that they form is spongy (or cancellous), but is later replaced by more compact bone tissue as concentric layers of bone are deposited on the inside surface of channels along the periosteum (outside of the bone). The channels thus become narrowed over time, forming canals through which lymph and blood vessels run. Once bone has stopped growing, the bone-forming cells occupy cavities in the bone, where they maintain the bone. The central cavity of long bones contains marrow (Lyman 1994: 72–78).

Antler is an outgrowth of bone and is structurally similar to long bone. The outer cortex is compact and bone-like, but the inner part is similar to cancellous bone, rather than marrow-filled, and the outer skin or velvet carries amino acids, minerals, proteins and a growth hormone to the growing antler. Following months of growth, antler ossifies and is shed annually after the breeding season. Thus, its deposit on a site does not require an animal's death.

Horns, on the other hand, are an outgrowth not of bone but of epitheleal tissue, like skin. Horn is made up of a fibrous

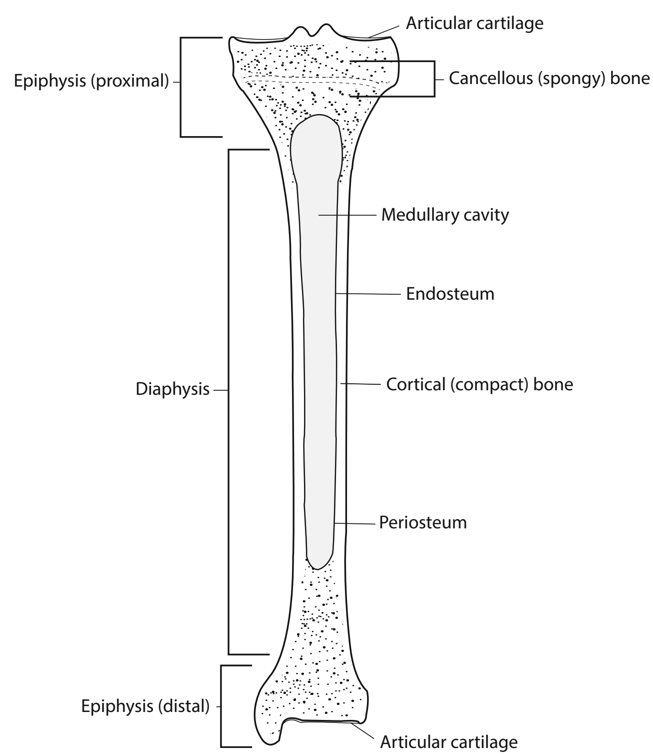


Fig. 15.4 Diagram of the parts of a generalized, mature mammalian long bone (W. Wadsworth)

material, called keratin, also found in hooves and claws. The horn grows by laying down layers of keratin over a horn core, which is a cancellous bony projection from the skull. Rhinoceros horns, however, do not have horn cores.

Although we tend to think of teeth as part of the skeleton, really they are part of the digestive tract and not bones at all. Teeth grow from tooth buds formed in the epithelial tissue that lines the mouth. They are composed of dentine, enamel and cementum, with a central nerve cavity. Elephant tusks consist almost entirely of dentine, with enamel only near the tips of young elephants' tusks.

15.3.1.1 Axial Parts

The axial skeleton of a vertebrate animal consists of the skull, vertebral column, ribs, and sternum. However, many zooarchaeologists conventionally treat the skull, and sometimes neck region, as separate portions of the skeleton (e.g., Stiner 1994: 242).

The skull has a cranium, which encases and protects the brain, facial bones that shape and protect the eyes, nose and mouth, and the mandible is the only really movable part of the skull, to accommodate eating and chewing. Sutures are immovable joints between the bones in the cranium, although they are still somewhat movable in very young animals. The dome (calvaria) of the cranium has a frontal bone, left and right parietal bones, left and right temporal bones, an occipital bone, sphenoid and ethmoid bone. The face includes two maxillae (upper mouth), a mandible (jaw), two zygomatic (cheek) bones, and two nasal (nose) bones, among others that are not visible. The foramen magnum is an opening that permits the spinal chord to enter the cranium. Near it, two bony projections called occipital condyles serve somewhat like hinges to allow the skull to nod up and down on the top of the vertebral column.

The vertebral column consists of a stack of somewhat cylindrical vertebrae, of which the foremost, articulating with the skull, are the atlas and axis. In a living animal, the vertebrae are separated by intervertebral disks of cartilage that protect the spinal chord, allow the spine to be flexible, and help support the animal's weight much like a shock absorber. The conventional segmentation of the vertebrae is among the cervical (neck), thoracic (upper or anterior back), lumbar (lower back), sacral (hip area) and caudal (tail) regions. Although they vary in detail, each vertebra has a body on which the disks rest, and a vertebral arch that encloses the spinal canal or foramen and is surrounded by "processes" that radiate from it. Tranverse processes are attachments for muscles and, in the case of thoracic vertebrae, for ribs. Openings in the vertebrae allow nerves to branch out from the spinal chord to various parts of the body.

The sternum and ribs are important for the protection of heart and lungs.

15.3.1.2 Appendicular Parts

The pectoral girdle consists of the scapula (shoulder blade), which serves as a broad attachment for muscles, and sometimes the clavicle (collar bone).

The forelimb consists of the humerus, radius, ulna, carpals, matacarpals, phalanges, and sometimes sesamoids.

The pelvic girdle (or innominate) consists of the left and right ilia, ischia, and pubis bones, all fused together and joined to the sacrum (fused sacral vertebrae) in the vertebral column. Males in some species also have a baculum (penis bone).

Each hindlimb consists of the femur (pl. femora), patella, tibia and fibula, tarsals, metatarsals, and phalanges.

15.3.1.3 Teeth

Mammals typically replace their deciduous (primary) teeth only once, so permanent teeth must serve through their adult lives. Mammal teeth include incisors, canines, premolars, and molars.

A mammal's tooth consists of the crown, exposed above the gum, and the root. The surface on the crown that makes contact with its opposite during chewing is called the occlusal surface, which itself has one or more projections called cusps. Enamel covers most of the crown in most species, but is absent, for example, on the lingual side of rodents' and rabbits' incisors. It is generally the cusps that are most helpful in identifying the animal to which a tooth belongs.

Some useful guides to mammalian skeletons include Beisaw (2013), France (2010), Fillios and Blake (2015), Hillson (2009), Huskey (2017), Schmid (1972), and Talmale (2017).

15.3.2 Skeletal Anatomy of Birds

Birds belong to the class Aves. Although their skeletal structure is similar to that of mammals, and they have many bones in common, avian skeletons also have unique features and ones they share with reptiles, especially the subclass Archosauria, which includes dinosaurs and crocodiles. Among these are features that allow them to run on their hind legs, making the forelimbs available for evolution into wings (Fig. 15.5). Pneumatization, meaning that many bones are hollow or contain air sacs to lighten avian skeletons (Bellairs and Jenkin 1960: 243, 289–293), is another flight-adapted feature.

The avian skull has several reptilian and other unusual features. In some birds, there is a craniofacial joint that allows the upper jaw, and not only the lower one, to move relative to the braincase. To accommodate their eyes, birds usually have extremely large orbits that rarely form full circles (King and McLelland 1984: 46). Instead, there is a ring of 10 to

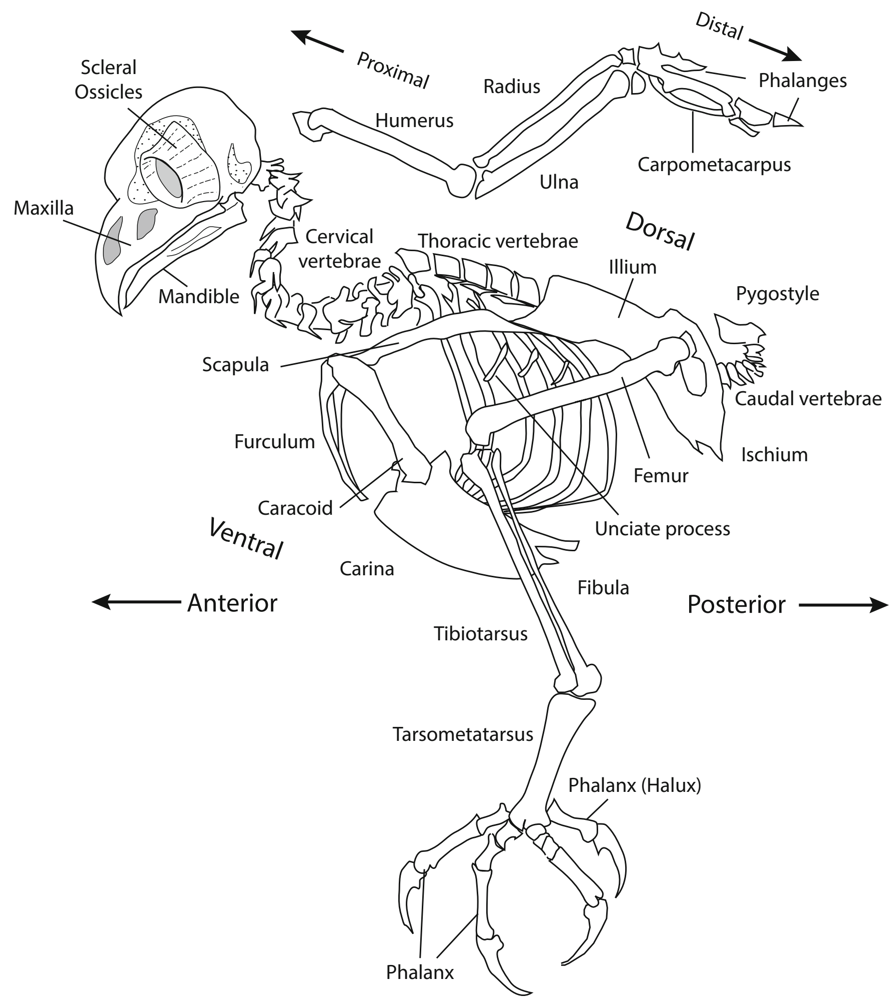


Fig. 15.5 Avian skeletal elements illustrated with an owl (*Bubo bubo*), with wings and one leg removed and, upper right, left wing of a swift (*Micropus apus*; W. Wadsworth after Bellairs and Jenkin 1960)

18 small, overlapping bones, *scleral ossicles*, to provide stiffer support for the eyeball (Bellairs and Jenkin 1960: 288). Bird's tongues actually have a skeleton, called the hyobranchial apparatus. This allows birds to probe into narrow spaces for food. With only one occipital condyle instead of two, birds can turn their heads almost full-circle.

In the axial skeleton, the number of vertebrae varies considerably but how many are in each portion of the column is often uncertain because vertebrae are often fused and it is difficult to distinguish cervical from thoracic vertebrae (King and McLelland 1984: 51). Fusion of parts of the vertebral column make the bird's trunk more rigid to support it in flight. The cervical vertebrae compensate by being more numerous and more mobile than in most mammals. Unlike most vertebrates, the articular surfaces of avian vertebrae are saddle-shaped, with the anterior surface concave in the tranverse but convex in the dorso-ventral plane, and are arranged in a way that allows the anterior part of the neck to move mainly forward and the middle part backward, so that the neck tends to be S-shaped (Bellairs and Jenkin 1960: 249; King and McLelland 1984: 52). All the cervical vertebrae except the atlas have ribs or their vestiges, mainly fused with the vertebrae, but sometimes the last one or two vertebrae have long and movable ribs articulated with them. The thoracic ribs, as in some reptiles, have uncinated processes, small bones at an angle to the ribs that provide extra muscle attachment. In flying birds and penguins, the sternum usually has a pronounced keel (carina) that strengthens it by providing a cross-section much like a steel girder and more area for pectoral muscle attachment, to provide strength for flying or swimming. Finally, there are often five to eight free caudal vertebrae and up to ten fused elements that form an upturned rump-post, or pygostyle, for moving some birds' tail feathers.

The pectoral girdle and forelimbs are usually adapted for flying. The girdle is strong to deal with compression stresses during flight, and the coracoids and clavicles (often fused into a furculum or "wishbone") act as struts that hold the wings away from the sternum. The scapula is usually long. A bird's wings are supported mainly by the "arm" bones — the humerus, radius and ulna — rather than the "hand" bones that support the wings of bats and pterosaurs (Bellairs and Jenkin 1960: 255). The proximal end of the humerus is flattened and has two prominent crests for muscle attachment. The ulna usually shows small knobs for attachment of quills of feathers. Fusion of distal carpals with three of the metacarpals creates the compound structure known as the *carpometacarpus*.

The pelvic girdle and hindlimbs are similar to those of some dinosaurs. The sutures between pelvic bones tend to disappear, while the illium is fused to the synsacrum, the whole thus becoming a rigid structure that carries the bird's weight when it is walking. In most birds, the ilium and ischium do not meet ventrally in a symphysis. This makes the pelvic outlet more open, allowing females to lay large, hard-shelled eggs. The acetabulum, into which the proximal femur fits, is completely perforated instead of a cup-like socket, and a facet on the ilium, above the acetabulum, prevents the femur from pushing through when there is weight upon it. The more distal parts of the hindlimb include the tibiotarsus (fusion of tibia with the proximal tarsal bones), with an anterior extension called the cnemial process, and fibula (usually very small in size), sometimes a patella, and a tarsometatarsus (fusion of distal tarsal bones with elongated metatarsals). Because the distal end of the last is made up of three fused metatarsals, it branches into three pulley-shaped processes called trochleas (Bellairs and Jenkin 1960): 263). Sometimes, as in fighting cocks, there is also a spur on the inner side of the tarsometatarsus. Birds vary considerably in the arrangement of their feet, most birds having four toes with three, four, and five phalanges, but others having three or even two.

There are several guides to avian skeletons (e.g., Cohen and Serjeantson 1991; Gilbert et al. 1996; Shufeldt 1909). Serjeantson (2009) provides a good review of the archaeology of birds.

15.3.3 Anatomy of Bony Fish

Unlike the skeletons of mammals, those of bony fish do not experience re-modelling or resorption; the bones simply keep growing throughout the fish's life. Like reptiles, fish continually replace their teeth because they also have a life-long supply of tooth buds. The rate of fish's bone growth varies seasonally, resulting in distinct growth rings that are helpful in determining the fish's age at death and even season of death. Fish differ considerably from mammals in skeletal anatomy, although many of the same classes of body part and terms for orientation still apply (Cannon 1987). The body parts of greatest importance to archaeologists are the otoliths (ear stones) and vertebrae, because these have greater probability of preservation than other elements. However, sometimes even scales are preserved and can be useful sources of data.

Otoliths are small concretions of calcium salts, principally calcium carbonate, that occur in a fish's inner ear. They appear to be part of its control over balance, or to help with depth perception or hearing (Casteel 1976: 17–18). The smallest otoliths are not very helpful for identification, but the larger ones, called statoliths, vary substantially in shape by species, sex and other factors, making them very useful for identification. They occur in three pairs: the sagittae, asterisci, and lapilli.

The outer face of most otoliths is flat or concave, sometimes with concentric rings or annuli, but with little relief, while the inner face is highly sculptured in ways that facilitate identification, sometimes even to species level (Fig. 15.6). Both the sagitta and the asteriscus tend to show a groove, called sulcus, extending roughly horizontally across the inner surface and often bordered by a ridge. Two protrusions, the anterior rostrum and the antirostrum, which juts out on the anterior dorsal edge, flank the anterior end of the sulcus. Various online databases of otolith shapes are very useful for identification of fish in different regions of the world (e.g., Fisheries and Oceans Canada n.d.; Sadighzadeh et al. 2012).

Fish vertebrae (Fig. 15.7) consist of a spool-shaped centrum to which a number of spines are attached. The concave

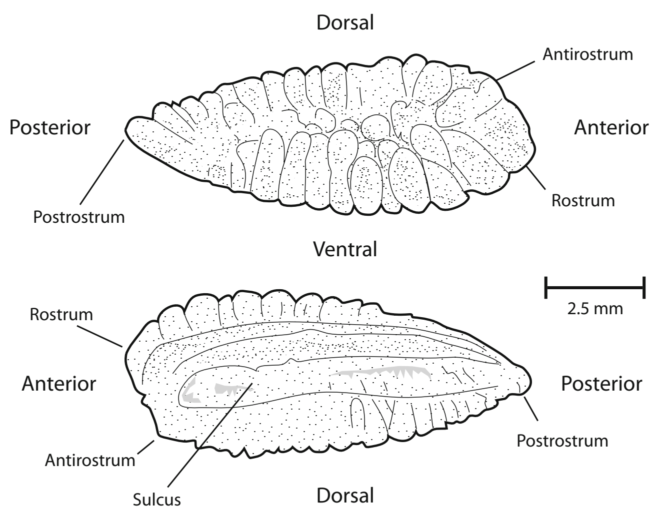


Fig. 15.6 Lateral (top) and medial (bottom) views of a sagittal otolith

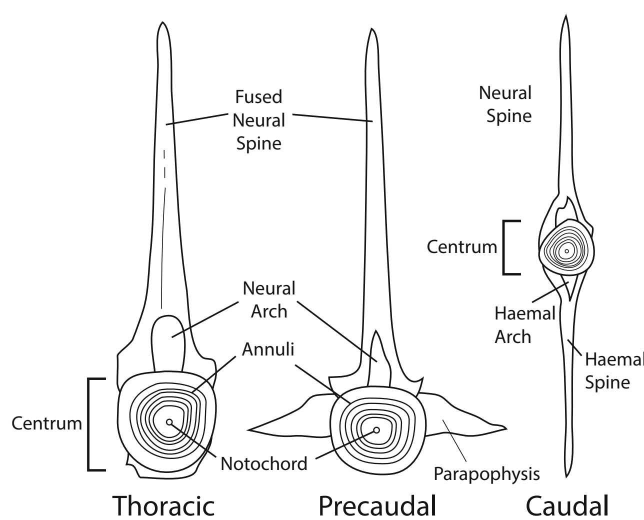


Fig. 15.7 Generalized anterior view of thoracic, precaudal and caudal vertebrae as found in some genera of bony fish

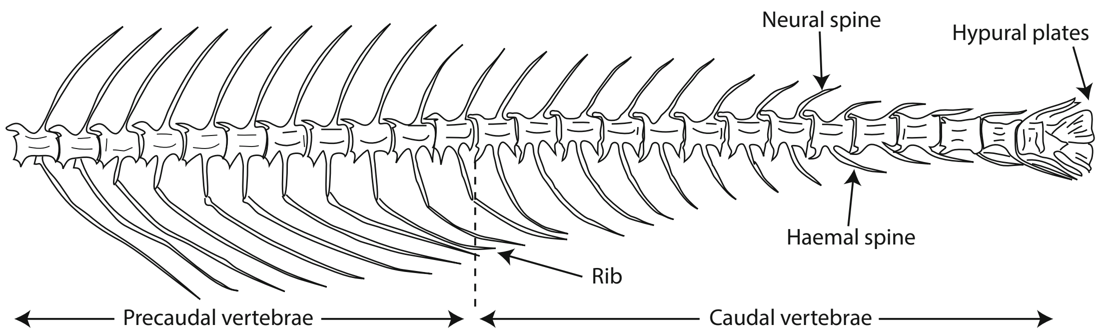


Fig. 15.8 Generalized lateral view of a bony fish's vertebral column, divided more generally into precaudal and caudal regions (after Cannon 1987: 21)

face or faces of centra show annuli, or growth rings, useful for ageing fish (Hofkamp and Butler 2017; Smith 1992). Casteel (1976: 76–78) recommends classifying fish vertebrae according to nine classes, extending posteriorly from the basioccipital, where the skull articulates with the spinal column. These are the proatlas, the atlas, the second vertebra, the Weberian vertebrae, the thoracic vertebrae, the precaudal vertebrae, the caudal vertebrae, the penultimate vertebra, and the ultimate vertebra. However, it is also common to use just two classes, precaudal and caudal, for fish vertebrae (Fig. 15.8). The spines are not always preserved in archaeological specimens but can include the neural spine on the dorsal side of the centrum, parapophyses or transverse processes that extend laterally from the centrum, and haemel spines (or ribs) flanking the centrum's ventral surface, which can sometimes be fused. The passage for the spinal cord is the neural arch. The ultimate vertebra has a urostyle at the posterior end where it begins the fanning out of the tail.

15.3.4 Anatomy of Molluscs

Mollusc shells constitute the only invertebrate remains that archaeologists routinely collect, although analysis of archaeological insect and even parasite remains can sometimes be very important (e.g., Buckland 2000; Buckland et al. 2004; Graham 1965; Morrow et al. 2016; Panagiotakopulu 2000; Yeh et al. 2016). Archaeological study of mollusc shells is called archaeomalacology (Bar-Yosef Mayer 2005).

The major classes of molluscs (phylum Mollusca) are the Monoplacophora, the Amphineura (including subclass Polyplacophora), the Gastropoda (including snails, slugs, limpets, and abalone), the Scaphopoda (including Dentalium or "tusk" shells), the Bivalvia or Pelecypoda ("bivalves" such as clams and oysters), and the Cephalopoda (including squid, octopus, nautilus, and extinct ammonites; Wilbur and Yonge 1964). Molluscs' hard tissue occurs, if at all, in the form of one or more exoskeletons, or shells, formed from deposited calcium carbonate with only a very small amount of protein, that serve as external protection for the animal's soft, unsegmented body (Waselkov 1987).

The shells occur in a number of major varieties. Gastropods, for example, can have an overall helicoid (a long cone wound around a shorter cone) shape, but are described as ventricose, if each whorl bulges outward between sutures, flat-sided, if the whorls are flattened, or turreted, if the upper part of each whorl projects outward below the suture. Other gastropods have shells shaped like simple cones. In bivalves, the two shells are hinged, each valve a near mirror-image of the other, and occur in a wide variety of symmetrical and asymmetrical shapes. There are several good introductions to shells globally and regionally (e.g., Claassen 1998; Wye 1991).

As you would expect, the terminology for the parts of mollusc shells is very different from that for skeletons.

Helicoid gastropod shells are oriented relative to the apex, vertically along the axis of the spiral, with apex at top (Fig. 15.9). The coiling is normally upwards and clockwise (destral) when viewed from the apex, but there are exceptions with sinistral coiling or with the spire extending downwards. In helicocones that lay new whorls that touch the central axis, this creates a solid or hollow central pillar, called the columella, only visible in a broken shell. In hollow columellae, the lower opening is called the inferior umbilicus. The protoconch is the original shell, secreted while the animal was embryonic or larval, at the innermost part of the shell's helicocone. The mouth, or aperture, is near the bottom of the shell and its growing edge, or peristome, includes an outer lip, or labrum, away from the axis of the shell and inner columellar lip or labium, close to the shell's axis. Where part of the peristome lies over the previous whorl, it is called the parietal lip. An important attribute for identification is the angle of the outer lip relative to the shell's axis. Other features of the aperture can include the siphonal or anterior canal, which allows the gastropod to suck water into its mantle cavity, and which can be an elongated tube in some species. At the adapical end of the aperture, there may be a posterior notch or canal, which facilitates the discharge of feces. The youngest (largest) whorl, forming one complete turn of the helicocone, is the body whorl; all other whorls combine to make the spire. The maximum diameter of the body whorl is known as its periphery. Tiny spiralling folds on the surface of the whorls are known as striae, while periodic thickenings of shell along the length of the helicocone are ribs, which appear to intersect the striae. Where the ribs appear parallel to the axis of the shell, they are orthocline ribs. If their adapical ends are turned toward the direction of growth, they are prosocline, and if turned in the opposite direction, they are opisthocline. At the microscopic level, there are also tiny growth lines, created by secretion along the peristome, that can be seasonal, allowing age estimates by counting the lines. In some species, there are long spines protruding from the ribs, or series of out-turned flanges, called varices (sg. varix), both of which originate in an out-turning of the outer lip. In some species, mature animals stop growth of the helicocone and instead secrete bosses (denticles or teeth) on the inside of the aperture. A septum (pl. septa) is an interior growth of shell, extending from the columella, to block off the uppermost whorls of the spine when the animal no longer occupies them (Fretter and Graham 1962: 50–66; Wye 1991: 16).

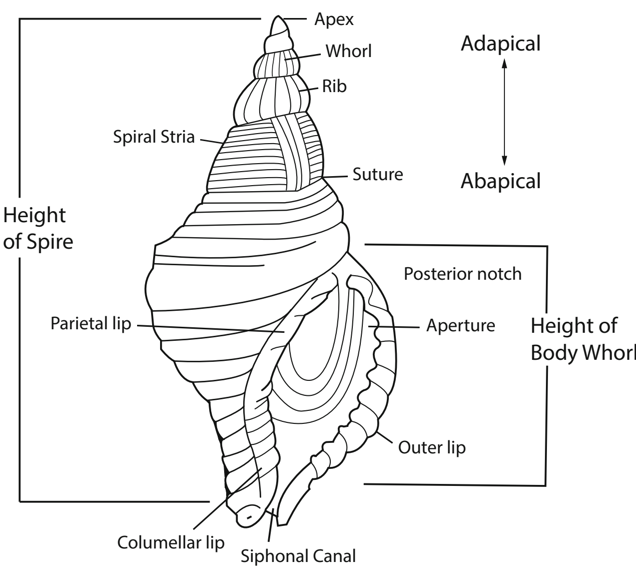


Fig. 15.9 Anatomy of a helicocone (W. Wadsworth)

Bivalve shells are oriented relative to the umbo (or beak) and nearby ligament (hinge), inner and outer surfaces, anterior and posterior ends, and left and right valves (Fig. 15.10). The shell of a bivalve grows from the margin of the mantle, as in gastropods, but the mantle is divided into two symmetrically arranged lobes, from which shell growth must be identical to ensure that the margins of the two valves meet when the shell is closed. The mantles' margins have three parallel folds, of which the outer one secretes the peristracum and outer calcareous layer of the shell, the middle one has sensory functions, and the inner one controls water flow into the mantle cavity. The shell itself has three major layers: the outer, horny periostracum, made of protein, the outer calcareous layer, consisting of calcium carbonate in a matrix of protein conchiolin, and the inner calcareous layer, which

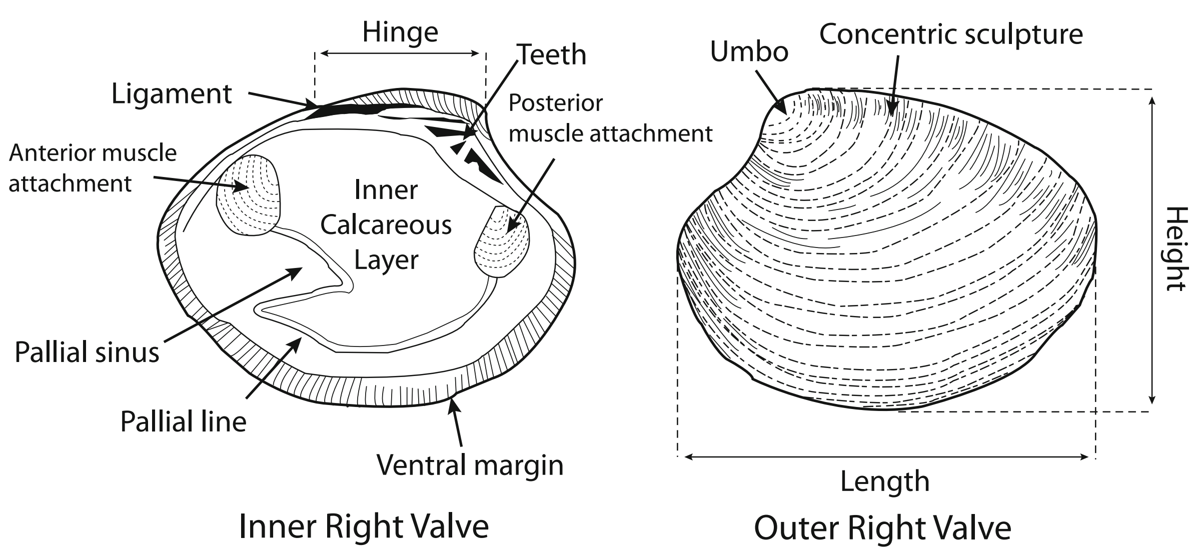


Fig. 15.10 Anatomy of bivalve shells illustrated by inner and outer right valve (W. Wadworth)

sometimes has the iridescent quality of mother-of-pearl. Along the bivalve's mid-line, an uncalcified, elastic ligament serves as a hinge. While the ligament is rarely preserved in archaeological assemblages, hinge "teeth," ridges and grooves on either side of the ligament are often important attributes for identification. Each tooth fits into its corresponding groove or socket on the opposing valve. The teeth ensure that the two valves line up properly when the shell is closed, and that they will not slip against one another (Yonge and Thompson 1976: 142–149).

As with fish, seasonal variations in temperature and food availability affect shell growth in many species of molluscs. For some, there may be no shell growth at all in winter (Davis 1987: 83–90). Molluscs accumulate shell when specialized parts of the mantle secrete an organic base, conchiolin, and inorganic salts of calcium around its growing edge. Later deposition of calcareous material thickens the shell over its whole inner surface. But shell deposition is not constant. Some bivalves, for example, only secrete shell when the animal is submerged and the shell is open. In intertidal bivalves, such as the cockle, there is a new increment of shell at each high tide. Hundreds of these small increments, separated by thin dark lines, accumulate during each growing season (generally summer) and are separated by grooves, or "growth cessation rings," when only extremely thin increments, or none at all, are added to the shell. This has potential for extremely precise seasonality assessments in some cases, and sclerochronology is the study of these accretionary growth lines (Burchell et al. 2013; West et al. 2018). Some gastropods also show seasonal increments, representing either winter cessation of growth due to cold, or summer cessation when gastropods in arid regions burrow to avoid dessication. These growth increments are important evidence for seasonality.

There are several useful guides to identification and analysis of molluscs (e.g. Allen 2017; Evans 1972; Preece 1981).

15.4 Identifying Faunal Remains

The first step in identification of mammalian, avian, and bony fish remains is to determine to what skeletal part a particular specimen belongs. Then, morphological idiosyncracies on the specimen allow us to narrow down the taxonomic possibilities to a particular family, genus or species. Molluscan identifications more often involve whole or fairly complete shells. Standard references are helpful for identifying more common animals (e.g., Beisaw 2013; Broughton and Miller 2016; Cohen and Serjeantson 1991; Cornwall 1964; Hillson 2009; Huskey 2017; Pales and Garcia 1971; Prehn et al. n.d.; Schmid 1972; Talmale 2017; Walker 1985). However, shapes of elements also vary among individuals for reasons other than interspecific differences, including developmental changes over the animal's life cycle, sexual dimorphism, pathologies (Baker and Brothwell 1980), and other factors. For some zooarchaeological research, identifying the animals' sex, age or season of death is as important as identifying their genus or species.

A criticial tool for identification is a reference collection of modern skeletal elements, teeth, shell and horn or antler from reliably identified specimens of known species, age, and sex, as determined before or shortly after the animals' deaths (Mori 1979). Ideally, it should include multiple specimens of each species to represent various ages and both sexes, as well as intra-species variation in size and shape. It is also useful for the documentation of specimens to show the place and date of each animal's killing or capture, its mass and other useful size measures (e.g., length or stature) at the time of death. Because most of the species of archaeological interest evolved only very slowly, modern specimens are usually a good guide to identification of most archaeological specimens. Most of the exceptions pertain to Pleistocene animals that are now extinct. For these, one can use a collection in a museum of natural history that is sufficiently complete. Generally, zooarchaeologists or the institutions in which they work build up their own reference collections (Fig. 15.11). This sometimes requires hunting, trapping, fishing, or finding cooperative zoos, park services, and butchers or slaughterhouses, making exchange arrangements with other reference collections, and sometimes collecting roadkills, sea animals washed up on beaches, or birds that collided with tall buildings. As a quality matter, it is essential to ensure that specimens are correctly identified and to minimize the possibility of specimens being returned to the wrong box in the collection (Gobalet 2001: 385).

Fig. 15.11 View of a zooarchaeological laboratory with a large reference collection (courtesy Howard Savage Zooarchaeological Laboratory, University of Toronto)

15.5 Preparing Skeletal Remains for a Reference Collection

Once carcasses are available, it is necessary to deflesh the animals to remove all traces of tissue that could attract pests while specimens are in storage (Anderson 1965; Casteel 1976: 8–16; deWet et al. 1990; Egerton 1968; Friedman 1973; Hangay and Dingley 1985: 326–65; Hildebrand 1968; Mayden and Wiley 1984; McDonald 2006), a particularly messy and smelly task that requires attention to health and safety (Chap. 10). A common method is a combination of defleshing manually with sharp tools, simmering the partly defleshed specimen in water that is hot but not boiling, often with some detergent, enzyme, sodium perborate or sodium hydroxide in the water (Chapman and Chapman 1969; Jakway et al. 1970; Ossian 1970; Simonsen et al. 2011), and maceration (allowing bacteria to rot the partly defleshed specimen in lukewarm water). This is fairly slow and is best accomplished under a fume hood to evacuate steam and very unpleasant odors. Safe disposal of water that contains rotted tissue and enzymes can also be a problem, since disposal down the drain may violate health-and-safety regulations or municipal laws. Another method is to use dermestid beetles (*Dermestes* spp.), moth larvae, or meal worms (larvae of the beetle, *Tenebrio molitor*) to clean the specimens (Borell 1938; Banta 1961; Grayson and Maser 1978; Allen and Neill 1950). This works very well but is even slower and beetle colonies require careful maintenance (Graves 2005), including preventing their escape into your collection, making this a practical option only for labs that have a rather constant flow of defleshing work. Yet another option, only practical at coastal laboratories, is to place eviscerated specimens in perforated containers in the intertidal zone of the sea, where marine isopods will deflesh them fairly quickly (Casteel 1976; Packard 1959).

It is also necessary to degrease bones. Among the ways to do this are soaking in ammonia, cooking in hot but not boiling solution of 50% Ammonia or a degreasing detergent (McDonald 2006; Mairs et al. 2004). You should not use older methods that involved carbon tetrachloride or the fuel from camp stoves as the former is a carcinogen and the latter is flammable, and both require disposal as hazardous waste. In some cases, complete degreasing may require burying the specimen outdoors in a pit filled with clean sand, in a well-drained area. If you do this, make sure the specimens are enclosed in a fine mesh so that you do not lose any elements, and that the location of the pit is well-marked so you can find it again. GPS coordinates are also helpful (McDonald 2006; Hoffmeister and Lee 1963).

Nested screens are necessary to capture the disarticulated remains that result from any of these methods. When everything is clean and dry, you will need to label each element carefully and accurately by species, age, sex, element (including left and right, where relevant), identification number, and preferably also the collection date or date of death. Store the labelled specimens safely and appropriately in a pest-free environment with stable temperature and moderate humidity (Chap. 9 and Williams 1999), where they are easily retrievable for reference. This may require thousands of plastic or other insect- and rodent-proof containers of various sizes, sturdy shelving, and adequate table or counter space.

Keep in mind that your choice of methods for processing faunal remains has implications both for the long-term stability of the collections and for the potential usefulness of the specimens for various kinds of anticipated and unanticipated research (Williams 1999).

15.6 Determining Sex

Some skeletal parts vary by sex in their size, shape, or both. For example, in equids (horses), males usually have large canine teeth, while females usually either lack canines or have only vestigial ones. In both cervids and bovids, females usually either lack horns or antlers or have ones that are different in shape from those on males. The pelvis frequently shows differences related to birthing. In most mammals, males' bones tend, on average, to be larger and more robust than the same bones in females of the same age, and males often have greater overall body size. There are exceptions, however; in hares, for example, does are larger than bucks. Castration of males, furthermore, slows epiphyseal fusion and allows more longitudinal growth. Sometimes, bones that support the extra weight of the male's large horns or antlers are good indicators of sex. Using overall size as an indirect measure of sex can work well in species with strong sexual dimorphism, such as goats, cattle, and seals, but in some species, the overlap between size distributions of males and females is too great for this to be effective. In addition, size varies with age as well as sex, so in is necessary to control for age by such indicators as fused ephiphyses (Klein and Cruz-Uribe 1984: 40).

15.7 Age at Death

There are several sources of evidence for determining an animal's age (Klein and Cruz-Uribe 1984; Morris 1972).

One of these is size, and ratio-scale measurements can be useful in this respect, but size also varies for a number of reasons other than age, including the animal's sex, dietary stress, domestication, and isolation on islands.

For mammals, a good indicator is fusion of epiphyses. A layer of cartilage separates epiphyses from the shaft of long bones in young animals. Later, ossification fuses the epiphyses to the shaft at predictable ages. The fact that, for some bones, the two ephyses fuse at different times allows even better precision in ageing well-preserved specimens. The schedule of fusion for modern animals of a species thus allows us to create a somewhat coarse scale for estimating age in months (Davis 1987: 39; Klein and Cruz-Uribe 1984: 43). However, unfused epiphyses are less likely to survive burial than fused ones.

Teeth provide a more precise indirect measure of age. As horse-traders have known for centuries, patterns of dental eruption and tooth wear are excellent clues to an animal's age (Brothwell 1989; Hillson 1986: 176–223). Deciduous teeth are replaced on a schedule that is known in modern animals, allowing us to age young animals very precisely if their mandibles or maxillae are preserved and show the stage of tooth eruption (Deniz and Payne 1982). Tooth wear is another indirect measure of age. Young animals show very little wear on the occlusal surfaces of teeth (apart from post-depositional abrasion), while old animals may show considerable wear. Although the degree of wear depends on a number of factors besides age, such as the amounts of grit and acid in diet, it is possible to make a scale (Fig. 15.12) that allows rough estimates of age (e.g., Klein and Cruz-Uribe 1984; Spinage 1973). In cases where animals' deaths were likely seasonal or catastrophic, certain teeth may cluster into groups on the basis of their wear, ones with almost no wear coming from animals in their first year, those with slightly more wear from animals 1 or 2 years old, and so on (Klein and Cruz-Uribe 1984: 45). Tooth eruption and wear are better indicators when whole rows of teeth are available.

In some herbivorous animals, with high-crowned teeth whose growth stops early in life, tooth wear results in ever-diminishing crown height, as measured from the top of the root. For example, Spinage (1972) shows an exponential decrease in crown height of zebra teeth.

Another ageing method involves counting growth increments (*annuli*) in the cementum of mammalian teeth (Benn 1974; Low and Cowan 1963; Morris 1972, 1978). Like tree rings in many tree species, these *annuli* provide an age record because deposition of cementum, the tissue that mineralizes collagen fiber on the roots, varies seasonally and continues throughout the animal's life. In stained thin sections under polarized light (Fancy 1980), the *annuli* show as alternating translucent and opaque bands that we can count to determine the animal's age at death. There is potential for error in that cementum may not deposit uniformly, that some of it may be resorbed or lost through abrasion, and "secondary bands" may be mistaken for annuli (Klein and Cruz-Uribe 1984: 44–45). In addition to age, this method can be used for estimating the season of death.

For fish, fairly constant bone growth over the fish's life allows size measures and ratios among them to work well for age estimates (Casteel 1976: 93–123). However, as with cementum in mammals, the *annuli* in otoliths, scales, and the centra of fish vertebrae (Fig. 15.7) can also be used as measures of age, as has been recognized for more than two centuries (Casteel 1976). Hofkamp and Butler (2017) note that counting annuli in radiographs of vertebral centra, however, provides biased measures of absolute age because, at least in salmon, the internal walls in the centra that impede passage of x-rays are not annual. They suggest that incremental growth on the faces of centra, which results from a different process than the growth of the walls and is visible in reflected light, may provide more accurate measures of age.

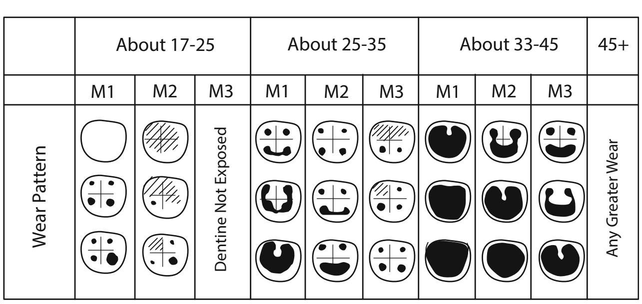


Fig. 15.12 Ordinal scale of molar wear for estimating the age of humans (Y. Salama, after Brothwell 1981: 72)

15.8 Zooarchaeological Sampling

Archaeological samples of animal remains often come from archaeologists' "hand-picking" of relatively large teeth, bones and their fragments during excavation, or from screening with sieves with meshes on the order of 3 mm in aperture. However, smaller remains from fish, rodents (or "micromammals"), birds, reptiles, and other small animals often come from wet-screening with small meshes or the flotation systems used to extract plant remains (see p. 274; Stewart 1991). Although zooarchaeological sampling begins in the field, it continues in the lab as analysts do not always have the resources to examine the excavated collections in their entirety. In both field and lab situations, it is important to identify the population being sampled (e.g., deposited assemblage or sample assemblage) and to take precautions to minimize the impact of factors that could contribute to biased estimates of abundance, diversity, representation of skeletal parts, or other measures (see Chap. 6).

15.9 Zooarchaeological Quantification and Interpretation

Once identifications are complete, there remains the question of how best to analyze the data. As already discussed in detail in Chap. 7, fragmentation and other taphonomic challenges influence the abundances of animal specimens. No single measure that zooarchaeologists have conceived can escape these problems. It is critically important to consider which measure is likely to perform better in the context of a specific research question and specific taphonomic circumstances (Lyman 2019).

Much of the confusion in zooarchaeological quantification stems from failure to appreciate the statistical shortcomings of Minimum Number of Individuals (MNI) as a general substitute for Actual Number of Individuals (ANI), and unwarranted concern over lack of independence between individual specimens on the grounds that they could come from the same animal. Among other problems, because MNI is a minimum, there is no way for it to yield accurate estimates of ratios among taxa except in the unusual case of nearly perfect preservation and nearly 100% sampling fraction of the deposited assemblage. It can, however, have a role in more targeted applications, such as examining the differential survival of skeletal parts (Lyman 1994). Meanwhile, the fact that multiple specimens contributing to Number of Identified Specimens (NISP) in a sample could come from a single animal is not the problem with NISP. Rather, the problem is that different individual animals and especially different taxa probably did not contribute to the sample proportionally. Not only do some animals have more bones and teeth than others in their skeletons, some animals' skeletal parts are much less likely to survive or be found by archaeologists. That is a taphonomic problem that none of our measures escape, so attention to the taphonomic environment and bone density and geometry is critical. Most importantly, selecting the right measure depends on the researcher's purpose.

One measure that is specific to zooarchaeology, Minimum Number of Elements (MNE), is designed to address one of these purposes: differential representation of skeletal parts. It in turn contributes to measures of economic utility that help us interpret the uses to which humans put the animals they hunted or herded.

15.9.1 Elemental Abundance and Choice of Cut

When humans and other bone accumulators kill, butcher, and consume an animal or use its parts for non-food purposes, they affect the distribution of bones and bone fragments in one or more deposited assemblages. They do this by selective transport of body parts from the kill site to the site of processing, consumption, or discard and by differential destruction of bones. Hunters may take the most valued cuts, such as upper limbs and pelvic girdle, back to a base camp while abandoning parts with low meat/bone ratios, such as skull, spine and lower limbs unless these latter have, or are attached to, something of value for a non-food purpose, such as fur or trophies (Binford 1978). In pastoral and farming societies, particular body parts can similarly be transported far from the location where butchering took place.

The Minimum Number of Elements (MNE) is a measure that zooarchaeologists use to evaluate this phenomenon. It is like the first step in calculating MNI (see pp. 129–132) — counting the total for each left and right element — but skips MNI's final step of using only the highest of these values. In other words, it is just the minimum number of animals that can account for the complete and fragmentary remains of a particular skeletal element or skeletal portion (e.g., forelimbs or thoracic portion of the axial skeleton). Comprehensive MNE (cMNE) is the same as MNE except that counting does not occur until after analysts have attempted to refit fragments as much as possible into complete elements (Bunn and Kroll 1986; de Ruiter 2004).

Aside from its use in studying differential use of body parts, MNE or cMNE is also useful for creating mortality or survivorship curves, which depend on particular elements, such as long-bone diaphysis and teeth, that are useful for ageing (see pp. 326–327).

An alternative to MNE is NDE (Number of Distinct Elements), which is the "number of times a diagnostic landmark is represented in a sample" of some skeletal element (Morin et al. 2017: 952). Summing these NDE values for each taxon provides measures of relative taxonomic abundance.

Archaeologists have attempted to interpret these choices in terms of "utility indices." Binford (1978: 72–74; 1981) introduced a General Untility Index (GUI) as a measure of each bone's value for meat, marrow and grease, and a Modified General Utility Index (MGUI) that makes allowances for the value of resources, like hides and fur, that might be attached to otherwise low-valued parts. Binford used these to distinguish between kill sites and camps and to infer whether hominids or other carnivores were responsible for bone accumulations.

Later researchers modified this approach by introducing a Meat Utility Index (MUI) (Lyman 1992; Lyman et al. 1992; Metcalfe and Jones 1988):

$$MUI = g - d$$

where g is the gross mass of the body part, including both meat and bone, and d is the dry mass of the bone only. This gives high scores to bones that would have carried a large mass of tissue, such as femora, even if the ratio g/d is relatively low and they are difficult to transport. Ringrose (1993: 146) proposes indices based on g/d, $(g\text{-}d)/d$, or $(g\text{-}d)/g$, with preference for the last. However, the decision to use MUI or one of Ringrose's alternatives should depend on whether your purpose is to examine the relative amounts of meat that an assemblage represents, or to explore hunters' decisions with regard to the "efficiencies" of balancing transport cost and the value of the body part (the so-called 'Schlepp effect').

While utility indices were originally for terrestrial mammals, subsequent research expanded their use to other classes of food animals, such as marine mammals (Lyman et al. 1992; Savelle and Friesen 1996).

Distributions of cut-marks on bones provide another type of evidence for butchering practices and choice of cut, although it is also necessary to distinguish butchering marks from other kinds of bone-surface modifications, such as carnivore damage. Zooarchaeologists have assessed bone-surface modifications in various ways, including just counting total cut-marks on a particular anatomical part (cut-mark count), cut-mark clustering, median cut-mark length, and statistical generalization of cut-mark distributions using Geographic Information Systems (GIS) software (Dominguez-Rodrigo 1997; Merritt 2016; Salladié et al. 2014).

One of the problems with bone utility measures, if not used with caution, is that differences in probability of bone preservation are confounding factors. Binford's MGUI, for example, is inversely correlated with bone density, one of the main factors in bone survival, so that patterns he describes as "bulk utility strategy" are also explicable by differences in probability of bone survival (Grayson 1989; Lyman 1992).

However, this problem encouraged research into variations in bone density and then density-mediated attrition, one of the major contributors to differential destruction of bone (e.g., Faith et al. 2007; Lam et al. 1998; Lyman 2013; Stiner 2002; and see pp. 309–310).

15.10 Paleoecology and Climate Change

Both archaeologists and paleontologists have long used animal remains as evidence for extinct ecologies and major climatic events (Butzer 1971: 258–262; Lyman 2017). For example, Davis (1977) reconstructs changes in the environment around the site of Kebara Cave in Israel on the basis of changes in the relative contributions of fallow deer and gazelle in the remains of animals that the cave occupants had hunted. Davis expected the deer to prefer wooded habitats and the gazelle to occupy more open landscapes. However, hunters' cultural and personal preferences and the potential for post-depositional alterations to the assemblages are confounding variables that present challenges to the validity of this approach. Despite these issues, the fact that faunal remains from archaeological sites provide a local record presents some advantages over some other records, such as pollen cores or ice cores, that typically come from distant locations or generalize climate at regional or global scales.

Rodents and other small animals (microfauna) on which owls prey may provide a more representative picture of the environment around a cave site (Talmale and Pradhan 2009; Tchernov 1968). Barn owls that live in caves may sample the rodents in their hunting range and, after consuming these animals, disgorge the bones and fur as "owl pellets." Changes in the relative frequencies of rodents that, in modern ecologies, prefer wetter or drier habitats can then contribute to reconstruction of changes in the owls' catchment area. However, a confounding factor is that owls may show preference for larger prey (Yom-Tov and Wool 1997). Microfauna are having an increasingly important role in paleoenvironmental reconstruction (e.g., Rhodes et al. 2018).

Marine shell from shell middens provides excellent evidence for environmental changes (West et al. 2018). Shell middens often have long sequences of use, while the shells can be analyzed for stable isotopes, especially oxygen isotopes whose ratios depend on the temperature of the water in which the molluscs lived and the water's ^{18}O content and salinity, the latter varying with evaporation and freshwater input from streams. Tracking variation in the oxygen isotopic ratios can provide very fine-grained information because of our ability to determine the seasonality of individual shells so precisely. This allows us to determine, for example, whether a climate change involved very seasonal events, such as el Niño or monsoons, rather than more general shifts in average precipitation or temperature. However, accurate matching of sequences from shell middens with those from other kinds of climate records requires radiocarbon dates either from associated animal bones or charcoal, or from the shells themselves, with correction for the marine reservoir effect — the difference between the abundance of ^{14}C in local seawater and the atmosphere (e.g., Jones et al. 2010; see Chap. 20).

Land snails can also be useful environmental indicators because many species have very distinct preferences for habitat, including water availability, soil pH, temperature, and calcium levels (Butzer 1971: 265–266; Evans 1972; Thomas 1985). Snails tend to occur in associations of several taxa, and grouping them into classes such as "shade-loving," "open country," "stream-edge," and "marshland" is helpful by avoiding too much emphasis on particular taxa (Thomas 1985: 140–144). The main drawback to using snail remains is that we cannot be certain of their association with the deposits in which they occur. Because snails burrow to escape inhospitable seasons, such as cold winters or hot, arid summers, snail remains can be considerably younger than the deposit in which they occur, and dating them is not easy because, like marine molluscs, they have substantial radiocarbon reservoir effects that are even more difficult to estimate. Comparison of terrestrial and marine shells from the same contexts helps to calculate more accurate reservoir corrections, however (Carvalho et al. 2015). Generally, it is better to have a long stratified sequence of snail remains so that we can detect likely environmental changes at the site whose timing may be imprecise, but whose general pattern is still strong.

Because they also may have very specific habitat requirements, insects can also be useful in this respect. Where conditions permit their preservation, parts of the insects' exoskeletons can serve for environmental reconstruction. Kenward (1985) shows how they can help us distinguish indoor from outdoor contexts, something particularly useful where architecture is not very substantial or where it is uncertain whether an architectural space is a room or a courtyard. Some archaeologists have also used the presence of exoskeletal parts from insects that infest grain stores to identify storage areas whether or not plant remains have been preserved (e.g., Kislev 1991; Panagiotakopulu and Buckland 1991). However, some potential sources of error are that wind, birds, and transport of building materials can introduce insect parts into deposits.

Strontium (Sr) and oxygen (O) isotopic ratios in the bones or tooth enamel of humans and other animals can be useful signatures for the geographical origins of those individuals. Strontium isotopes vary with the geological characteristics of the land where crops or grazing lands were growing, and animals who eat those plants retain those signatures in their tooth enamel. Oxygen isotopes similarly vary with the source of drinking water. In combination, these two isotopic signatures can identify geographical origins to facilitate studies of human migration and hunting ranges (e.g., Evans et al. 2012; Gregoricka 2013), or interregional movements of herd animals (Arnold et al. 2016).

15.11 Seasonality

Faunal remains are among the best indicators of the season of use or occupation at archaeological sites.

Where the fauna include large numbers of cervid (deer) remains, the fact that these animals shed antlers annually can be a source of seasonality data. However, because humans often curate antler to use as tool material presents challenges to the validity of this evidence.

The presence of migratory birds among the fauna can also serve to show that a site was occupied during the season when you would expect these birds to be present in the region. As Muñiz (1998) shows, however, birds' phenology (behavior with respect to climate and season) can change in response to climate change, food availability, and other factors. Bird species whose modern breeding grounds were glaciated during the Pleistocene, for example, must once have had very different migration patterns than they do today. Evidence from migratory birds can still be very helpful if sufficiently abundant and used with appropriate caution.

Marine molluscs provide particularly good evidence for seasonality at shell middens and other sites that contain substantial amounts of marine shell (Deith 1985). Molluscs increase steadily in size as they age, but, where seasonal environmental variation is pronounced, increments of shell growth alternate with periods of almost no shell growth in winter or summer. For quahogs, for example, Quitmyer et al. (1985) could identify six seasonal phases depending on the thickness and translucence of the outer increment.

For mammals, tooth-eruption can sometimes provide evidence for season as well as age at death, but only if the animals had a short and predictable birthing season and the sample size of sets of teeth is large. Of more practical use are those *annuli* in the tooth cementum. The number of annuli indicates age and the outermost increment, unless it is abraded away, will provide evidence of the season of death (Bourque et al. 1978; Lieberman and Meadow 1992; Miracle and O'Brien 1998). More recent improvements to the measurement of dental cementum, including automated analysis of variation in brightness of pixels in digital images, have greatly increased its accuracy, reliability and speed over older methods (Greenfield et al. 2015).

15.12 Diet and Food Preferences

Zooarchaeologists often try to reconstruct past dietary preferences, including the relative contributions of various taxa or age groups to the "menu." But they are also interested in food practices, such as butchering, marrow extraction, cooking, and meat storage. More broadly, this is one aspect of the archaeology of food (Hastorf 2017), which of course also includes plant-based resources.

One topic that has attracted considerable archaeological attention is change in hunters-gatherers' apparent behaviors from a fairly narrow niche (hunting mainly one or two animal taxa) to a broad niche (hunting and collecting a wide range of food resources). We could measure diet breadth in terms of overall diversity (Kelly 2007), such as with the Expected Species index, $s(m)$(see Chap. 7), or by comparing the diversity of taxa selected with that of taxa available in the environment (Feinsinger et al. 1981). A broadening of the niche could be associated with a prelude to domestication, where it has sometimes been called the "broad-spectrum revolution" (Flannery 1969; Zeder 2012c), but may also be related to changes in food availability as a result of climate change, population growth, overhunting, or some other factor. One interesting research focus is on changes in the contributions of small, slow-moving prey animals, such as tortoises or molluscs, relative to fast-moving ones, such as birds, hares and rabbits, or larger, fast-moving animals, such as gazelle, that vary in their procurement costs and meat weight (Stiner and Munro 2002; Stiner et al. 2008). However, attempts to interpret these changes sometimes give insufficient attention to human agency in their reliance on climatic factors (see Zeder 2012c).

In other cases, archaeologists are just interested in food choices for their own sake or, for example, to explore the role of food in social relationships (e.g., Twiss 2012). These provide a good example of the need to select valid indirect measures of food quantities (see Chap. 7), since the various animals that can contribute to diet vary considerably in the amount of meat and other possible food resources (marrow, fat, offal) they can provide by species, sex and age. Other confounding factors include the possibility that hunters only brought selected animal parts back to the site, that they acquired some animals or their parts for non-food purposes, or that they disposed of some kinds of animal remains off-site (e.g., Binford 1978).

Today, archaeologists can also make inferences about people's diet, from both animal and plant resources, on the basis of the carbon and nitrogen isotopes in their bones. Shifts in the ratios of nitrogen isotopes are related to animals' place in the food chain, or **trophic level**. The ratio of $^{15}N/^{14}N$ isotopes (or $\delta^{15}N$) tends to be higher in carnivores than in herbivores, often by a factor of about 3‰. Carbon isotopes from bone are also important. As carbon is more relevant to the plant part of diet, it will be discussed in more detail in Chap. 16, but it is also helpful to show both carbon and nitrogen isotopic ratios on the same graph (Fig. 15.13). These isotopic signatures in human bone have been particularly revealing in coastal regions where people often ate fish or sea mammals that were themselves at a fairly high trophic level (e.g., Santana-Sagredo et al. 2016). However, the pathways that cause ^{15}N enrichment are poorly known, and the interpretation of nitrogen-isotopic evidence is not as straightforward as once assumed (Hedges and Reynard 2007).

Lipids, including fatty acids and their derivatives, have been a particularly valuable source of evidence for past diets (Evershed 2008; Malainey 2011: 201–218). Pottery vessels used to cook animal or plant foods, or both, absorb some of these lipids into their fabric, while some may also adhere to interior surfaces. We can extract the degraded lipids from the pottery by drilling the ceramic and using solvents, sometimes with the aid of microwave processing (Evershed et al. 2002; Gregg and Slater 2010). The residues are decanted and filtered, different compounds separated by further solvents, and fatty acids converted into methyl esters for analysis (see p. 204).

Lipid residues have featured in attempts to discover the origins of dairy products (e.g., Dudd and Evershed 1998; Gregg et al. 2009) and to distinguish various animal fats that could have figured in prehistoric diets. Lipids are characterized by the number of carbon atoms and the number of double-bonds in their fatty acid chain, so that $C_{16:0}$, for example, means a chain with 16 carbon atoms and no double bonds (palmitic acid). In Fig. 15.14, we see peaks in a gas chromatography spectrum for $C_{16:0}$, $C_{18:2}$, $C_{18:1}$, and $C_{18:0}$ (palmitic, lineolic, oleic, and stearic acids). Since organically-sourced carbon is one of the main constituents of these molecules, we can also look for evidence of carbon

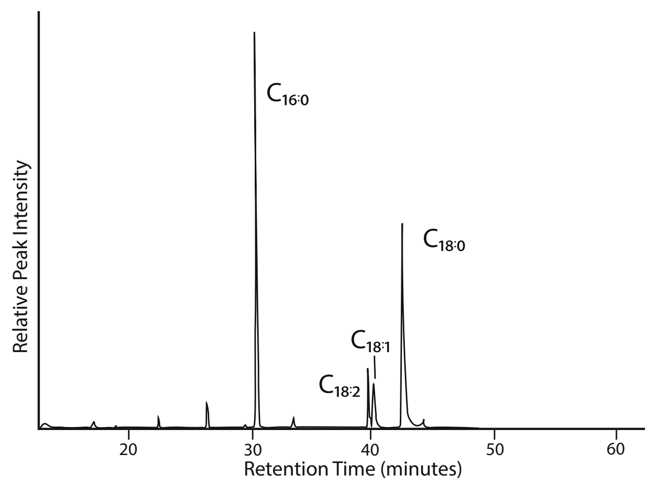


Fig. 15.14 Portion of a gas-chromatogram of fatty acids extracted from a Neolithic potsherd at al-Basatîn, Jordan, displaying peaks for fatty acids $C_{16:0}$, $C_{18:0}$, $C_{18:1}$, and $C_{18:2}$, (Y. Salama, after Gregg et al. 2009: 942). Because each fatty acid has different chemical properties, they exit the column of the gas chromatograph at different times, called retention times (x-axis), and the ratios among the peak heights for these fatty acids helps to identify animal and vegetable fats

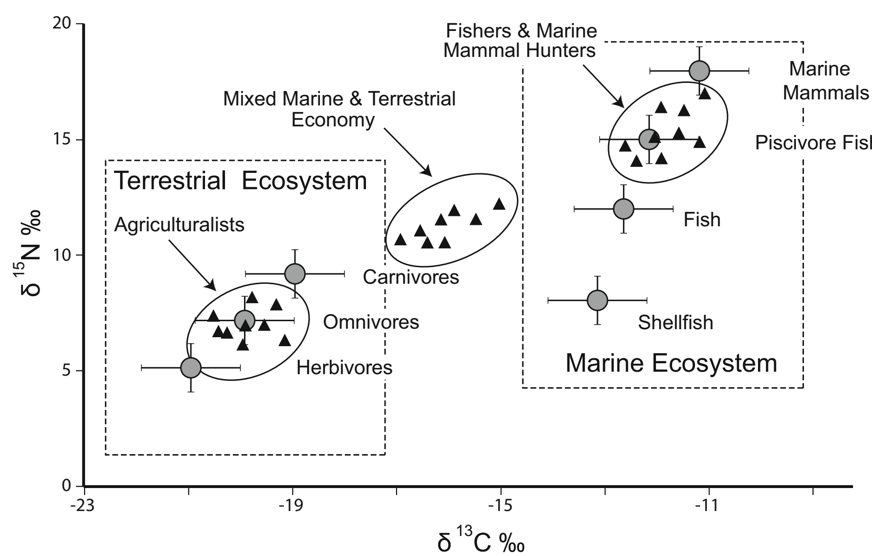


Fig. 15.13 Scatterplot of bone stable Nitrogen and Carbon ratios for Northwest Europe, with groupings for terrestrial and marine animals (Y. Salama, after Müldner 2009: 330)

fractionation due to the photosynthetic pathways of the plants that animals consumed (see p. 284).

15.13 Domestication of Animals

A major archaeological research area is the origins of food production and animals domesticated for either food or other purposes. Although archaeologists are particularly interested in the human behaviors involved in the shift from a hunting-and-gathering to a food-producing economy, they nonetheless rely a good deal on physical, evolutionary changes in animals' morphology and, increasingly, genetic evidence to understand this process.

Domestication is the process whereby human influences on the reproductive success of animals with varying characteristics — selection — causes the population of animals under human control, over time, to have physical or behavioral characteristics that the humans consider desirable. Technically, it is a mutualistic relationship between species, as the domesticated animal has greater reproductive success than it would have otherwise, while the humans benefit from the resources, services or companionship that the animal provides. It was the intentional selection by animal breeders to create or maintain breeds of livestock that gave Darwin the idea for natural selection (Darwin 1868). Domestication should not be confused with taming or herd management. Humans may have influenced the behaviors of certain animals for a long time before they became domesticated.

In the Old World, most of the interest has been in early domestications of dogs, goats, sheep, cattle and pigs, while, in the New World, it has been on dogs, guinea pigs and camelids. Some domesticated animals, such as Old World camels and chickens, were domesticated millennia after other domestic animals were already widespread.

Morphological evidence for domestication of animals can include size changes, shortening of the face, crowding of teeth, and loss of horns or changes in their shape (Bökönyi 1989). However, some of these changes may not have occurred until well after humans began to influence the animals' behaviors, and particularly reproductive behaviors. In addition, size changes may result from changes in prey selection, such as targeting younger animals or overhunting instead of domestication (Zeder 2012a, b, 2015).

15.14 Archaeogenetics and ZooMS

Today, molecular evidence such as ancient DNA is making rapid advances, not only to our understanding of animal domestication, but notably our understanding of hominin evolution (Malainey 2011: 237–253). Molecular evidence preserved in bones and teeth not only allows us to check on the accuracy of identifications of species and sex, ancient DNA (aDNA) is beginning to have revolutionary impacts on our understanding of the domestication of a broad range of animals, including dogs (Francis et al. 2011; Ovodov et al. 2011), goats (Bar-Gal et al. 2010; Daly et al. 2018), sheep (Demirci et al. 2013), cattle (Edwards et al. 2007) and pigs (e.g., Larson et al. 2007). Zooarchaeology by Mass Spectrometrey (ZooMS) is a method that allows us to identify genera, even in otherwise unidentifiable bone fragments, by analyzing proteins and peptides — chains of amino acides — in their bone collagen (Buckley et al. 2009).

Among the challenges to aDNA research is the fact that DNA degrades over time, especially in hot climates, and may undergo post-mortem mutations (Elsner et al. 2015; Vives et al. 2008). Even under ideal conditions, it is rarely preserved longer than a few hundred thousand years. However, certain elements have enhanced chances for DNA preservation, including teeth and the petrous part of temporal bones (Pinhasi et al. 2015).

When bones or teeth do preserve some DNA, it can be extracted, amplified and "sequenced" with Polymerase Chain Reaction (PCR). New-Generation Sequencing (NGS), a recent set of improvements to sequencing methods, allows faster throughput, lowers the cost, and provides better results by simultaneously sequencing millions of DNA fragments and reassembling them into longer chains. To avoid contamination with modern DNA, archaeogenetic labs employ extremely careful protocols, but also have species-specific PCR primers that reduce the risk of contamination by not amplifying any DNA fragments that do not belong to the species of interest (Knapp and Hofreiter 2010; Matisoo-Smith 2018; Shapiro and Hofreiter 2012; Rohland and Hofreiter 2007). In the case of aDNA from human remains, there are also ethical challenges (Prendergast and Sawchuk 2018).

While NGS is bringing down its cost, aDNA is still expensive, and this cost has encouraged use of ZooMS as a less costly alternative to identify species of bone fragments (Buckley 2018; Collins et al. 2010). This method also benefits from the fact that proteins, in most archaeological environments, have higher probability of preservation than DNA and can survive in large amounts for tens of thousands of years, with no requirement for amplification. After extracting the collagen proteins from powdered bone, they are reduced into peptides (shorter amino-acid chains). A mass spectrometer is used to measure the masses of these peptides to yield a peptide mass "fingerprint" that is highly distinctive. Principle Components Analysis (PCA) can separate the samples extremely clearly, allowing us to distinguish even closely related genera, such as sheep and goats, that are typically difficult to distinguish by conventional morphological analyses.

15.15 Secondary Products

Humans do not use animals only as a food source, or even of hides, bone, horn, antler or shell. Live animals can provide hair, wool, milk, traction and transportation. Sherratt (1981) coined the term, "secondary products revolution," to describe the innovation of keeping livestock to produce resources that do not require the death of the animal, with particular emphasis on dairy products, wool, and traction, the last especially for pulling plows. Sherratt argued that, in the Old World, this revolution took place substantially later than the Neolithic period, but more recent evidence suggests that some elements of this revolution, in some regions, may have occurred earlier than the Chalcolithic period, and in stages rather than all at once (Greenfield 2010). McCorriston (1997) goes on to specify a "fiber revolution" that included intensification of production of textiles made from both animal and plant fibers, and its implications for the alienation of women's labor.

Among the evidence that archaeologists examine to identify exploitation for secondary products are chemical residues in pottery (for dairying), bit wear on equid teeth (for horse-riding), and mortality profiles (see case study: Culling of Animals and Herding Strategies).

Case Study

Culling of Animals and Herding Strategies

Payne (1973) introduced the analysis of age-at-death patterns to infer whether an assemblage of sheep or goat bones resulted from a meat-producing, dairying, or wool-producing herding strategy. On the basis of ethnographic data, Payne reasoned that herders aiming to maximize productivity for meat would tend to kill male animals when they reach their optimum weight gain, keeping very few for breeding. For optimal milk production, most males would be killed off even younger, while, for wool production, males not needed for breeding can just be castrated, so there is less difference between the mortality profiles of males and females, and adults of both sexes tend to be killed somewhat younger, as the quality of their wool declines. Payne displayed "ideal" models for these culling strategies as mortality or survivorship profiles (Fig. 15.15), which are a kind of cumulative frequency curve that starts at 100% and gradually declines monotonically (i.e., can only go down) to 0% as animals are killed off. An important aspect of these models is the difference between treatments of male and female animals, but their use in conjunction with archaeological data, where the ageing of animal bones is coarse and sex is often impossible to determine, often makes it necessary to make less precise comparisons between archaeological distributions and mortality curves that combine males and females.

Others have used or adapted this method (e.g., Discamps and Costamagno 2015), including Vigne and Helmer's (2007) use of finer distinctions between two varieties of meat and two of dairy strategies. However, Marom and Bar-Oz (2009) question the reliability of these distinctions and test the culling models themselves, without distinguishing males from females, to see if they are statistically distinguishable, even aside from the problems of precision and accuracy of archaeological assessments

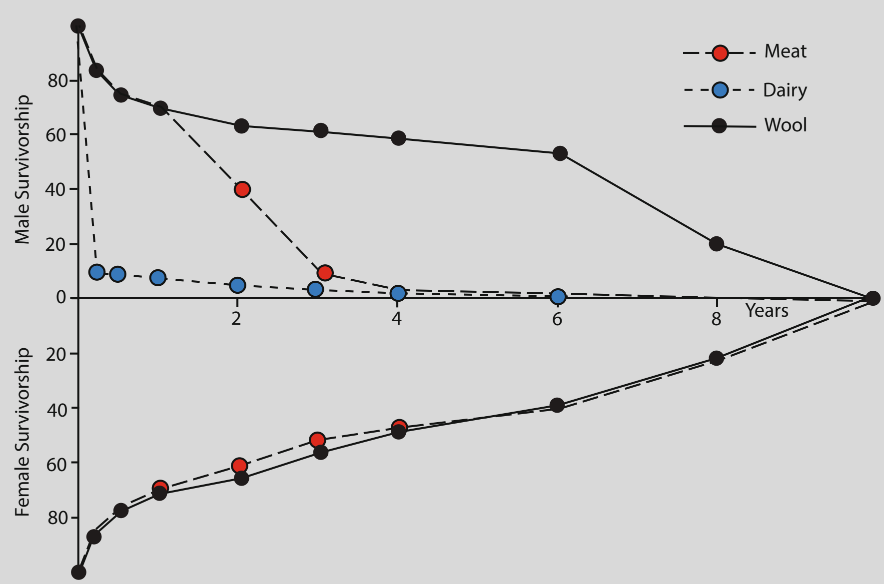


Fig. 15.15 Survivorship profiles in percent for male and female sheep or goats under ideal strategies for production of meat, dairy and wool (Y. Salama, after Payne 1973). Note that the curves for females differ little or not at all

(continued)

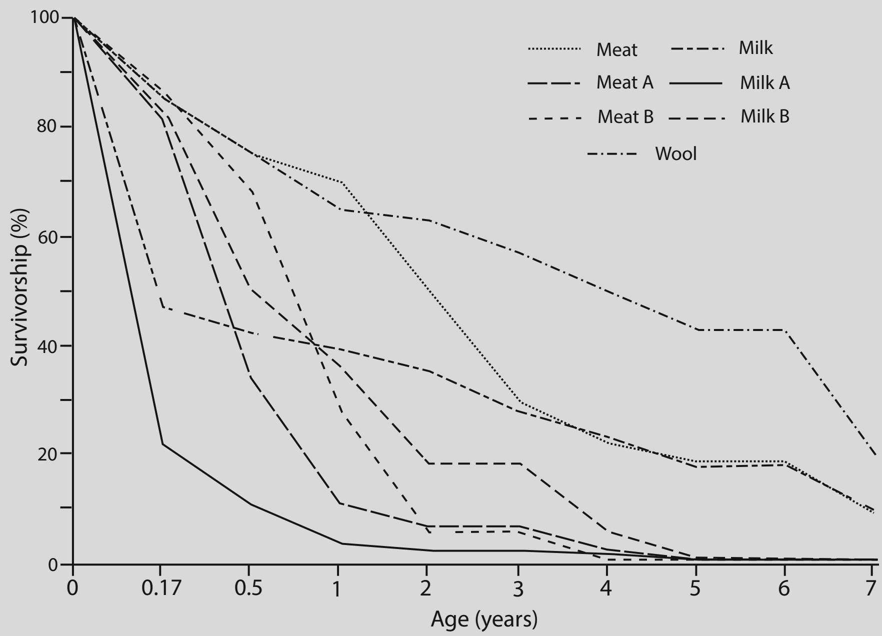


Fig. 15.16 Survivorship for sheep or goats, regardless of sex, under different herding strategies (Y. Salama, after Marom and Bar-Oz 2009). NOTE: the x-axis is not a linear scale, and the graph only shows a subset of the herding strategies

(Fig. 15.16). They used a test called the Kolmogorov-Smirnov test, which is appropriate for cumulative frequency distributions, and found that most of the models were statistically indistinguishable at 95% confidence, with only Vigne and Helmer's "Meat A" and "Meat B" models being well distinguished from one another, although not from other models. They suggest that simpler models may be more reasonable. A very high proportion (~0.8) of immature animals is highly suggestive of dairying, while equal representation of immature, subadult and adult animals suggests meat production, and a majority of adults is consistent with wool production. They also suggest use of fusion in distal metapodials and distal humeri for this purpose, as fusion of these elements is in consecutive years and they are sexually dimorphic (allowing distinction between culling of males and females) and generally well preserved.

Their analysis suggests need for caution in the more general use of archaeological mortality profiles to infer herding strategies (see also Brochier 2013; Price et al. 2016). Possibly this is an arena in which Bayesian methods would be helpful, by suggesting which strategies were more probable than alternatives, rather than taking a classical null-hypothesis-testing approach or just interpreting the curves visually and subjectively.

15.16 Inter- and Intra-Observer Differences in Faunal Identification and Measurement

Zooarchaeologists have long expressed concerns about the quality of their data, noting that agreement on identifications, even among experts, is not as strong as one would like (Driver 1992; Wolverton 2013). For example, Gobalet (2001) analyzed identifications and counts by five analysts, including himself, with doctorates in either Anthropology, Zoology, or Wildlife and Fisheries, of fish vertebrae, teeth and scales from a site in California. These analysts differed widely, with some identifying much larger numbers than others, or many more taxa than others. There was also much variation in the level of taxonomic specificity, with one analyst identifying many specimens to species level, two consistently identifying to family or higher levels, while three analysts sometimes identified to family but had only three families in common. As Gobalet (2001: 378) points out, the impacts of these uncertainties even on measures of taxonomic diversity, let alone abundance or ubiquity, are "daunting."

Lau and Whitcher Kansa (2018) express concern over interanalyst variation in the current trend toward "Big Data" and re-use of archived digital data. They agree that there is great potential for these projects, which take advantage of large datasets from multiple projects to study such large-scale processes as the spread of early domesticated animals, cultural variations in the consumption of animal taxa, and environmental and climatic influences on prey availability or hunting choices. However, it is also likely that lack of comparability in identification of taxa or methods of quantification will have negative impacts on the validity of results. They recommend incorporating quality-assessment protocols for inter-analyst variation into projects and including sufficient meta-data with published databases to allow informed re-use of the data in other applications.

Lyman and VanPool's (2009) evaluation of inter- and intra-analyst variation was not on taxonomic identification, but on measurements that the two of them made on astraguli of bighorn sheep (*Ovis canadensis*). Lyman measured two dimensions, maximum lateral length and distal breadth, of the sample of more than 60 astragali on two different occasions, separated by one month, which allows assessment of intra-observer variation on these dimensions. On the second occasion, he also measured two more dimensions, medial length and lateral depth, to permit comparisons with an independent analysis of the same collection of astagali by Lawler. Linear correlations (Fig. 15.17), technical error of measurement (TEM, see p. 9), relative TEM and coefficient of reliability (*R*) served to make these comparisons.

For the comparisons of measurements that Lyman made and then repeated a month later, agreement was rather good, with relative TEM of 0.13 to 0.71 percent and *R* of 0.979 to 0.997 and no evidence of bias. However, there was somewhat poorer agreement between Lyman and Lawler, with relative TEM ranging from 0.81 to 2.1 percent, and *R* of 0.775 to 0.892 but, once again, no evidence of bias. Lyman and VanPool (2009: 500) point out that it is much easier to check on reliability through re-measurement than by simple comparisons of datasets that could, and probably do, have different population parameters. While they do not describe it as such, they are essentially recommending lot acceptance sampling (e.g., MIL-STD-1916 1996; see also p. 207–208), a protocol whereby expert measurers check "lots" of measurements at specified intervals.

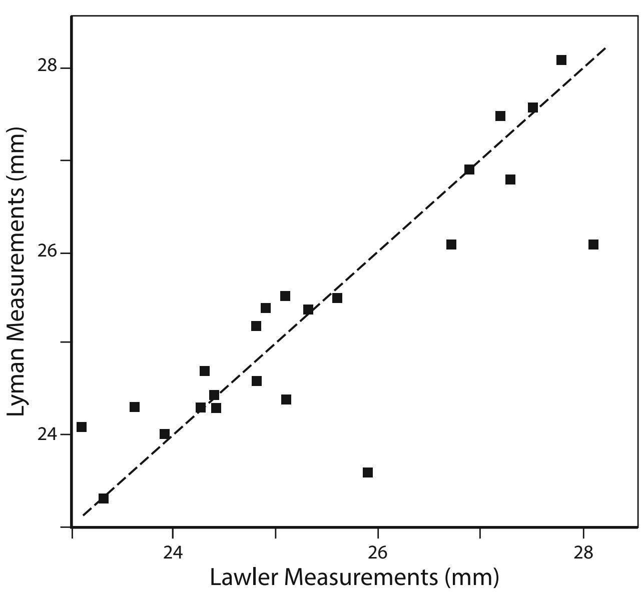


Fig. 15.17 Example of a correlation between Lyman's and Lawler's measurements of the distal breadth of 23 bighorn sheep astragali (after Lyman & VanPool 2009: 495). The dashed diagonal line represents perfect agreement. This was the case with the poorest agreement between the two analysts (TEM = 0.53 mm, relative TEM = 2.1%, R = 0.775)

15.17 Summary

- Faunal remains are very diverse, ranging from bones, teeth, antler and horn, through otoliths and shells, to microscopic spherulites, chemical residues, and isotopic ratios
- Taphonomic processes are a constant concern for zooarchaeologists, as differential preservation can have huge impacts on the representation of taxa and body parts
- Along with paleoethnobotanists, zooarchaeologists distinguish among five different kinds of populations from the life assemblage, though several levels of degradation and selection, to the sample assemblage
- Identification of faunal remains depends on knowledge of basic zoological taxonomy and functional anatomy, but also relies heavily on the availability of reference collections
- Estimating a specimen's age at death ranges from fairly coarse ordinal measures with large margins of error, such as tooth eruption and wear or fusion of epiphyses, to very precise measures based on annual and sometimes even seasonal growth increments, as in fish vertebrae and otoliths
- A key aspect of zooarchaeological interpretation is the selection of an appropriate quantitative measure in the context of a research question and specific taphonomic circumstances

References Cited

Albarella, U., & Trentacoste, A. (Eds.). (2011). *Ethnozooarchaeology: The present and past of human-animal relationships*. Oxford: Oxbow Books.

Allen, M. J. (Ed.). (2017). *Molluscs in archaeology: Methods, approaches and applications*. Oxford: Oxbow Books.

Allen, E. R., & Neill, W. T. (1950). Cleaning mammal skeletons with meal worms. *Journal of Mammalogy, 31*(4), 464.

Anderson, R. M. (1965). *Methods of collecting and preserving vertebrate animals* (National Museum of Canada bulletin 69). Ottawa: National Museum of Canada.

Arnold, E. V., Hartman, G., Greenfield, H. J., Shai, I., Babcock, L., & Maeir, A. M. (2016). Isotopic evidence for early trade in animals between Old Kingdom Egypt and Canaan. *PLoS One, 11*(6), 1–12.

Baker, J., & Brothwell, D. (1980). *Animal diseases in archaeology*. London: Academic.

Banta, B. H. (1961). The use of clothes moth larvae (Lepidoptera; Tineidae) to prepare osteological specimens, with an annotated bibliography of the use of other arthropods for vertebrate skeletal preparation. *Wasmann Journal of Biology, 19*(2), 265–268.

Bar-Gal, G. K., Khalaily, H., Marder, O., Ducos, P., & Kolska Horwitz, L. (2010). Ancient DNA evidence for the transition from wild to domestic status in Neolithic goats: A case study from the site of Abu Gosh, Israel. *Ancient Biomolecules, 4*(1), 9–17.

Bar-Yosef Mayer, D. (Ed.). (2005). *Archaeomalacology: Molluscs in former environments of human behaviour*. Oxford: Oxbow Books.

Beisaw, A. M. (2013). *Indentifying and interpreting animal bones: A manual*. College Station: Texas A&M University Press.

Bellairs, A. d'A., & Jenkin, C. R. (1960). The skeleton of birds. In J. Marshall (Ed.), *Biology and comparative physiology of birds* (pp. 241–300). New York: Academic.

Benn, D. W. (1974). Annuli in the dental cementum of white-tailed deer from archaeological contexts. *Wisconsin Archeologist, 55*, 90–98.

Binford, L. R. (1978). *Nunamiut ethnoarchaeology*. New York: Academic.

Binford, L. R. (1981). *Bones: Ancient men and modern myths*. New York: Academic.

Borell, A. E. (1938). Cleaning small collections of skulls and skeletons with dermestid beetles. *Journal of Mammalogy, 19*(1), 102–103.

Bourque, B. J., Morris, K., & Speiss, A. (1978). Determining season of death of mammal teeth from archaeological sites: A new sectioning technique. *Science, 149*, 530–531.

Brochier, J. É. (2013). The use and abuse of culling profiles in recent zooarchaeological studies: Some methodological comments on 'frequency correction' and its consequences. *Journal of Archaeological Science, 40*, 1416–1420.

Brothwell, D. R. (1981). *Digging up bones* (3rd ed.). Oxford: Oxford University Press.

Brothwell, D. R. (1989). The relationship of tooth wear to aging. In M. Y. Iscan (Ed.), *Age markers in the human skeleton* (pp. 306–316). Splringfield: Charles Thomas.

Broughton, J. M., & Miller, S. D. (2016). *Zooarchaeology and field ecology: A photographic atlas*. Salt Lake City: University of Utah Press.

Buckland, P. (2000). *An introduction to palaeoentomology in archaeology and the BUGS database management system*. Umeå: Institutionen för arkeologi och samiska studier, Umeå universitet.

Buckland, P. C., Palagiotakopulu, E., & Buckland, P. I. (2004). Fossil insects and the Neolithic: Methods and potential. *Antaeus, 27*, 235–252.

Buckley, M. (2018). Zooarchaeology by Mass Spectrometry (ZooMS): Collagen fingerprinting for the species identification of archaeological bone fragments. In C. M. Giovas & M. J. LeFebvre (Eds.), *Zooarchaeology in practice: Case studies in methodology and interpretation in archaeofaunal analysis* (pp. 227–247). New York: Springer.

Buckley, M., Collins, M., Thomas-Oates, J., & Wilson, J. C. (2009). Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry, 23*, 3843–3854.

Bunn, H. T., & Kroll, E. M. (1986). Systematic buthery by Plio/Pleistocene hominids at Olduvai Gorge, Tanzania. *Current Anthropology, 27*, 431–452.

Burchell, M., Cannon, A., Hallmann, N., Schwarcz, H. P., & Schöne, B. R. (2013). Refining estimates for the season of shellfish collection on the Pacific Northwest Coast: Applying high-resolution stable oxygen isotope analysis and sclerochronology. *Archaeometry, 55*, 258–276.

Butzer, K. W. (1971). *Environment and archaeology. An ecological approach to prehistory* (2nd ed.). Chicago: Aldine.

Bökönyi, S. (1989). Definitions of animal domestication. In J. Clutton-Brock (Ed.), *The walking larder: Patterns of domestication, pastoralism and predation* (pp. 22–27). London: Unwin-Hyman.

Cannon, D. Y. (1987). *Marine fish osteology: A manual for archaeologists*. Burnaby: Archaeology Press.

Carvalho, C., Macario, K., De Oliveira, M. I., Oliveira, F., Chanca, I., Alves, E., et al. (2015). Potential use of archaeological snail shells for the calculation of local marine reservoir effect. *Radiocarbon, 57* (3), 459–467.

Casteel, R. W. (1976). *Fish remains in archaeology and paleoenvironmental studies*. London: Academic.

Chapman, D. I., & Chapman, N. (1969). The use of sodium perborate ($NaBO_3.4H_2O$) in the preparation of mammalian skeletons. *Journal of Zoology, 159*, 522–523.

Clark, J., & Kietzke, K. K. (1967). Paleoecology of the lower nodular zone, Brule formation, in the big Badlands of South Dakota. *Fieldiana Geology Memoir, 5*, 111–137.

Claassen, C. (1998). *Shells*. Cambridge: Cambridge University Press.

Cohen, A., & Serjeantson, D. (1991). *Manual for the identification of bird bones from archaeological sites*. London: Archetype Publications.

Collins, B. (2010). Element survivability of *Salmo salar*. *Journal of Taphonomy, 8*, 291–300.

Collins, B., & Steele, T. E. (2017). An often overlooked resource: Ostrich (Struthio spp.) eggshell in the archaeological record. *Journal of Archaeological Science: Reports, 13*, 121–131.

Collins, M., Buckley, M., Grundy, H. H., Thomas-Oates, J., Wilson, J., & van Doorn, N. (2010). ZooMS: The collagen barcode and fingerprints. *Spectroscopy Europe, 22*(2), 6.

Cornwall, I. W. (1964). *Bones for the archaeologist*. New York: The Macmillan Company.

Daly, K. G., Delser, P. M., Mullin, V. E., Scheu, A., Mattiangeli, V., Teasdale, M. D., et al. (2018). Ancient goat genomes reveal mosaic domestication in the Fertile Crescent. *Science, 361*(6397), 85–88.

Darwin, C. (1868). *The variation of animals and plants under domestication*. London: John Murray.

Davis, S. J. M. (1987). *The archaeology of animals*. London: B. T. Batsford.

Davis, S. J. M. (1977). The ungulate remains from Kebara cave. *Eretz Israel, 13*, 150–163.

de Ruiter, D. (2004). Relative abundance, skeletal part representation and accumulating agents of macromammals at Swartkrans. In C. Brain (Ed.), *Swartkrans: A cave's chronicle of early man* (pp. 265–278). Pretoria: Transvaal Museum Monographs.

Deith, M. R. (1985). Seasonality from shells: An evaluation of two techniques for seasonal dating of marine molluscs. In N. R. J. Fieller, D. Gilbertson, & N. G. A. Ralph (Eds.), *Palaeobiological investigations: Research design, methods and data analysis* (pp. 119–130). (BAR international series 266)). Oxford: British Archaeological Reports.

Demirci, S. K., Bastanlar, E. K., Dagtas, N. D., Piskin, E., Engin, A., Özer, F., et al. (2013). Mitochondrial DNA diversity of modern, ancient and wild sheep (Ovis gmelinii anatolica) from Turkey: New insights on the evolutionary history of sheep. *PLoS One, 8* (12), e81952.

Deniz, E., & Payne, S. (1982). Eruption and wear in the mandibular dentition as a guide to ageing Turkish angora goats. In B. Wilson, C. Grigson, & S. Payne (Eds.), *Ageing and sexing animal bones from archaeological sites* (pp. 155–205). (BAR British series 109)). Oxford: British Archaeological Reports.

deWet, E., Robertson, P., & Plug, I. (1990). Some techniques for cleaning and degreasing bones and a method for evaluating the long-term effects of these techniques. In E. M. Herholdt (Ed.), *Natural history collections: Their management and value* (pp. 37–47). Pretoria: Transvaal Museum.

Discamps, E., & Costamagno, S. (2015). Improving mortality profile analysis in zooarchaeology: A revised zoning for ternary diagrams. *Journal of Archaeological Science, 58*, 62–76.

Dominguez-Rodrigo, M. (1997). Meat-eating by early hominins at the FLK 22 Zinjanthropus site, Olduvai Gorge (Tanzania): An experimental approach using cut-mark data. *Journal of Human Evolution, 33*, 669–690.

Driver, J. C. (1992). Identification, classification and zooarchaeology. *Circa, 9*, 35–47.

Dudd, S. N., & Evershed, R. P. (1998). Direct demonstration of milk as an element of archaeological economies. *Science, 282*(5393), 1478–1481.

Edwards, C. J., Bollongino, R., Sheu, A., Chamerlain, A., et al. (2007). Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society B: Biological Sciences, 274*(1616), 1377–1385.

Efremov, I. A. (1940). Taphonomy: A new branch of paleontology. *Pan-American Geologist, 74*, 81–93.

Egerton, C. P. (1968). Method for the preparation and preservation of articulated skeletons. *Turtox News, 46*(5), 156–157.

Elsner, J., Schibler, J., Hofreiter, M., & Schlumbaum, A. (2015). Burial condition is the most important factor for mtDNA PCR amplification success in Palaeollithic equid remains from the Alpine foreland. *Archaeological and Anthropological Sciences, 7*(4), 505–515.

Evans, J. G. (1972). *Land snails in archaeology*. London: Seminar Press.

Evans, J. A., Chenery, C. A., & Montgomery, J. (2012). A summary of strontium and oxygen isotope variation in archaeological human tooth enamel excavated from Britain. *Journal of Analytical Atomic Spectrometry, 5*, 754–764.

Evershed, R. P. (2008). Organic residue analysis in archaeology: The archaeological biomarker revolution. *Archaeometry, 50*(6), 895–924.

Evershed, R. P., Dudd, S. N., Copley, M. S., & Mukherjee, A. (2002). Identification of animal fats via compound specific $\delta^{13}C$ values of individual fatty acids: Assessments of results for reference fats and lipid extracts of archaeological pottery vessels. *Documenta Praehistorica, 29*, 73–96.

Faith, J. T., Marean, C. W., & Behrensmeyer, A. K. (2007). Carnivore competition, bone destruction, and bone density. *Journal of Archaeological Science, 34*(12), 2025–2034.

Fancy, S. G. (1980). Preparation of mammalian teeth for age determination by cementum layers: A review. *Wildlife Society Bulletin, 8*(3), 242–248.

Feinsinger, P., Spears, E. E., & Poole, R. W. (1981). A simple measure of niche breadth. *Ecology, 62*(1), 27–32.

Fillios, M., & Blake, N. (2015). *Animal bones in Australian archaeology: A field guide to common native and introduced species*. Sydney: Sydney University Press.

Fisheries and Oceans Canada. (n.d.). Otolith shape, reference collections. www.dfo-mpo-gc.ca/science/aah-saa/otoliths/scientists/shape-eng.html. Accessed 30 May 2019.

Flannery, K. V. (1969). Origins and ecological effects of early domestication in Iran and the Near East. In P. J. Ucko & G. W. Dimbleby (Eds.), *The domestication and exploitation of plants and animals* (pp. 73–100). London: Duckworth.

France, D. L. (2010). *Human and nonhuman bone identification: A concise field guide*. Boca Raton: CRC Press.

Francis, G., Quignon, P., Hitte, C., & André, C. (2011). Toward understanding dog evolutionary and domestication history. *Biologies, 334* (3), 190–196.

Friedman, E. (1973). Preparation of faunal specimens. *American Antiquity, 38*(1), 113–114.

Gifford, D. P. (1981). Taphonomy and paleoecology: A critical review of archaeology's sister disciplines. Advances in archaeological method and theory, 4, 365–438.

Gilbert, M. B., Martin, L. D., & Savage, H. G. (1996). *Avian osteology*. Columbia: Missouri Archaeological Society.

Giovas, C. M., & LeFebvre, M. J. (Eds.). (2018). *Zooarchaeology in practice: Case studies in methodology and interpretation in archaeofaunal analysis*. New York: Springer.

Gobalet, K. W. (2001). A critique of faunal analysis; inconsistency among experts in blind tests. *Journal of Archaeological Science, 28*, 377–386.

Graham, S. A. (1965). Entomology: An aid in archaeological studies. *Memoirs of the Society for American Archaeology, 19*, 167–174.

Graves, R. C. (2005). *Beetles & bones: Care, feeding, and use of dermestid beetles*. South Berwick: Jillett Publications.

Grayson, D. K. (1989). Bone transport, bone destruction, and reverse utility curves. *Journal of Archaeological Science, 16*(6), 643–652.

Grayson, D. K., & Maser, C. (1978). An efficient, inexpensive dermestid colony for skeleton preparation. *Journal of Field Archaeology, 5*, 246–247.

Greenfield, H. J. (2010). The secondary products revolution: The past, the present and the future. *World Archaeology, 42*(1), 29–54.

Greenfield, H. J., Moore, N. C., & Steppan, K. (2015). Estimating the age- and season-of-death for wild equids: A comparison of techniques utilising a sample from the Late Neolithic site of Bad Buchau-Dullenried, Germany. *Open Quaternary, 1*(3), 1–28.

Gregg, M. W., & Slater, G. F. (2010). A new method for extraction, isolation and transesterification of free fatty acids from archaeological pottery. *Archaeometry, 52*(5), 833–854.

Gregg, M. W., Banning, E. B., Gibbs, K., & Slater, G. F. (2009). Subsistence practices and pottery use in Neolithic Jordan: Molecular and isotopic evidence. *Journal of Archaeological Science, 36*, 937–946.

Gregoricka, L. A. (2013). Residential mobility and social identity in the periphery: Strontium isotope analysis of archaeological tooth enamel from southeastern Arabia. *Journal of Archaeological Science, 40*(1), 452–464.

Hastorf, C. A. (2017). *The social archaeology of food: Thinking about eating from prehistory to the present*. Cambridge: Cambridge University Press.

Hangay, G., & Dingley, M. (1985). *Biological museum methods*. Orlando, FL: Academic.

Hedges, R. E. M., & Reynard, L. M. (2007). Nitrogen isotopes and the trophic level of humans in archaeology. *Journal of Archaeological Science, 34*, 1240–1251.

Hildebrand, M. (1968). *Anatomical preparations*. Berkeley: University of California Press.

Hillson, S. (2009). *Mammal bones and teeth: An introductory guide to methods of identification*. Walnut Creek: Left Coast Press.

Hoffmeister, D. F., & Lee, M. R. (1963). Cleaning mammalian skulls with ammonia hydroxide. *Journal of Mammalogy, 44*(2), 283–284.

Hofkamp, A. R., & Butler, V. L. (2017). On the validity of the radiographic method for determining age of ancient salmon. *Journal of Archaeological Science: Reports, 12*, 449–456.

Holtzman, R. C. (1979). Maximum likelihood estimation of fossil assemblage composition. *Paleobiology, 5*(2), 77–89.

Huskey, S. (2017). *The skeleton revealed: An illustrated tour of the vertebrates*. Baltimore: Johns Hopkins University Press.

Jakway, G. E., Raskin, W., & Thyle, T. (1970). Sodium perborate process for preparation of skeletons. *Turtox News, 48*(2), 65–67.

Jones, K. B., Hodgins, G. W. L., Etayo-Cadavid, M. F., Andrus, C. F. T., & Sandweiss, D. H. (2010). Centuries of marine radiocarbon reservoir age variation within archaeological Mesodesma donacium shells from southern Peru. *Radiocarbon, 52*, 1207–1214.

Keepax, C. A. (1981). Avian egg-shell from archaeological sites. *Journal of Archaeological Science, 8*(4), 315–335.

Kelly, R. (2007). *The foraging spectrum: Diversity in hunter-gatherer lifeways*. New York: Percheron Press.

Kendall, C., Høier Eriksen, A. M., Kontopoulos, I., Matthew, B., Collins, J., & Turner-Walker, G. (2018). Diagenesis of archaeological bone and tooth. *Palaeogeography, Palaeoclimatology, Palaeoecology, 491*, 21–37.

Kenward, H. K. (1985). Outdoors-indoors? The outdoor component of archaeological insect assemblages. In N. R. J. Fieller, D. Gilbertson, & N. G. A. Ralph (Eds.), *Palaeobiological investigations: Research design, methods and data analysis* (pp. 97–104). (BAR international series 266)). Oxford: British Archaeological Reports.

King, A. S., & McLelland, J. (1984). *Birds: Their structure and function* (2nd ed.). London: Baillière Tindall.

Kislev, M. E. (1991). Archaeobotany and storage archaeoentomology. In J. Renfrew (Ed.), *New light on ancient farming* (pp. 121–136). Edinburgh: Edinburgh University Press.

Klein, R. G., & Kruz-Uribe, K. (1984). *The analysis of animal bones from archaeological sites* (Prehistoric archeology and ecology series). Chicago: University of Chicago Press.

Knapp, M., & Hofreiter, M. (2010). Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes, 1* (2), 227–243.

Lam, Y. M., Chen, X., Marean, C. W., & Frey, C. J. (1998). Bone density and long bone representation in archaeological faunas: Comparing results from CT and photon densitometry. *Journal of Archaeological Science, 25*(6), 559–570.

Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., et al. (2007). Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences, 104*(39), 15276–15281.

Lau, H., & Whitcher Kansa, S. (2018). Zooarchaeology in the era of big data: Contending with interanalyst variation and best practices for contextualizing data for informed reuse. *Journal of Archaeological Science, 95*, 33–39.

Lieberman, D. E., & Meadow, R. H. (1992). The biology of cementum increments (with an archaeological application). *Mammal Review, 22*(2), 57–77.

Low, W. A., & Cowan, I. M. (1963). Age determination of deer by annular structure of dental cementum. *The Journal of Wildlife Management, 27*(3), 466–471.

Lyman, R. L. (1984). Bone density and differential survivorship of fossil classes. *Journal of Anthropological Archaeology, 3*(4), 259–299.

Lyman, R. L. (1992). Anatomical considerations of utility curves in zooarchaeology. *Journal of Archaeological Science, 19*(1), 7–22.

Lyman, R. L. (1994). *Vertebrate taphonomy*. Cambridge: Cambridge University Press.

Lyman, R. L. (2008). *Quantitative paleozoology*. Cambridge: Cambridge University Press.

Lyman, R. L. (2013). Bone density and bone attrition. In J. Pokines & S. A. Symes (Eds.), *Manual of forensic taphonomy* (pp. 22–43). Boca Raton: CRC Press.

Lyman, R. L. (2017). Paleoenvironmental reconstruction from faunal remains: Ecological basics and analytical assumptions. *Journal of Archaeological Research, 25*, 315–371.

Lyman, R. L. (2019). A critical review of four efforts to resurrect MNI in zooarchaeology. *Journal of Archaeological Science, 26*, 52–87.

Lyman, R. L., & VanPool, T. L. (2009). Metric data in archaeology: A study of intra-analyst and inter-analyst variation. *American Antiquity, 74*(3), 485–504.

Lyman, R. L., Savelle, J. M., & Whitridge, P. (1992). Derivation and application of a meat utility index for Phocid seals. *Journal of Archaeological Science, 19*, 531–555.

Mairs, S., Swift, B., & Rutty, G. N. (2004). Detergent: An alternative approach to traditional bone cleaning methods for forensic practice. *The American Journal of Forensic Medicine and Pathology, 25*(4), 276–284.

Malainey, M. E. (2011). *A consumer's guide to archaeological science*. New York: Springer.

Marom, N., & Bar-Oz, G. (2009). Culling profiles: The indeterminacy of archaeozoological data to survivorship curve modelling of sheep and goat herd maintenance strategies. *Journal of Archaeological Science, 36*(5), 1184–1187.

Matisoo-Smith, E. (2018). Ancient DNA in zooarchaeology: New methods, new questions and settling old debates in Pacific commensal studies. In C. M. Giovas & M. J. LeFebvre (Eds.), *Zooarchaeology in practice: Case studies in methodology and interpretation in archaeofaunal analysis* (pp. 209–225). New York: Springer.

Mayden, R. L., & Wiley, E. O. (1984). A method of preparing disarticulated skeletons of small fishes. *Copeia, 1984*(1), 230–232.

McCorriston, J. (1997). The fiber revolution: Textile extensification, alienation, and social stratification in ancient Mesopotamia. *Current Anthropology, 38*(4), 517–535.

McDonald, H. G. (2006). Vertebrate skeletons: Preparation and storage. *National Park Service Conserve O Gram, 11*(7), 1–8.

Meadow, R. H. (1980). Animal bones: Problems for the archaeologist together with some possible solutions. *Paléorient, 6*(1), 65–77.

Merritt, S. R. (2016). Cut mark cluster geometry and equifinality in replicated Early Stone Age butchery. *International Journal of Osteoarchaeology, 26*, 585–598.

Metcalfe, D., & Jones, K. T. (1988). A reconsideration of animal body-part utility indices. *American Antiquity, 53*(3), 486–504.

MIL-STD-1916. (1996). *Preferred methods for acceptance of product*. Washington, DC: Department of Defense.

Miracle, P. T., & O'Brien, C. J. (1998). Seasonality of resource use and site occupation at Badanj, Bosnia-Herzegovina: Subsistence stress in an increasingly seasonal environment? In T. R. Rocek & O. Bar-Yosef (Eds.), *Seasonality and sedentism: Archaeological perspectives from old and New World sites* (pp. 41–74). (Peabody Museum bulletin 6)). Cambridge, MA: Harvard University.

Moclán, A., Dominguez-Rodrigo, M., & Yravedra, J. (2019). Classifying agency in bone breakage: An experimental analysis of fracture planes to differentiate between hominin and carnivore dynamic and static loading using machine learning (ML) algorithms. *Archaeological and Anthropological Sciences, 11*(9), 4663–4680.

Mori, J. L. (1979). Procedures for establishing a faunal collection to aid in archaeological analysis. *American Antiquity, 35*(3), 387–389.

Morin, E., Ready, E., Boileau, A., Beauval, C., & Courmont, M. P. (2017). Problems of identification and quantification in archaeozoological analysis, part II: Presentation of an alternative counting method. *Journal of Archaeological Method and Theory, 24*, 938–973.

Morris, P. (1972). A review of mammalian age determination methods. *Mammal Review, 2*(3), 69–104.

Morrow, J. J., Newby, J., Piombino-Mascali, D., & Reinhard, K. J. (2016). Taphonomic considerations for the analysis of parasites in archaeological materials. *International Journal of Paleopathology, 13*, 56–64.

Müldner, G. (2009). Investigating medieval diet and society by stable isotope analysis of human bone. In R. Gilchrist & A. Reynolds (Eds.), *Reflections: 50 years of medieval archaeology* (pp. 327–346). Leeds: Maney.

Muñiz, A. M. (1998). The mobile faunas: Reliable seasonal indicators for archaeozoologists? In T. R. Rocek & O. Bar-Yosef (Eds.), *Seasonality and sedentism: Archaeological perspectives from old and New World sites* (pp. 25–39). (Peabody Museum bulletin 6)). Cambridge, MA: Harvard University.

O'Connor, T. (2008). *The archaeology of animal bones* (2nd ed.). College Station, TX: Texas A&M University Press.

Orton, D. C. (2012). Taphonomy and interpretation: An analytical framework for social zooarchaeology. *International Journal of Osteoarchaeology, 22*, 320–337. https://doi.org/10.1002/oa.1212.

Ossian, C. R. (1970). Preparation of disarticulated skeletons using enzyme-based laundry presoakers. *Copeia, 1970*(1), 199–200.

Ovodov, N. D., Crockford, S., Kuzmin, Y., Higham, T. F. G., Hodgins, G. W., van der Plicht, J., et al. (2011). A 33,000-year-old incipient dog from the Altai Mountains of Siberia: Evidence of the earliest domestication disrupted by the last glacial maximum. *PLoS One, 6* (7), e22821.

Packard, A. (1959). Preparation of skeletons by marine animals. *Tuatara, 7*(3), 119–120.

Pales, C., & Garcia, M. A. (1971). *Atlas osteologique pour server à l'identification des mammifères du Quaternaire*. Paris: CNRS Editions.

Panagiotakopulu, E. (2000). *Archaeology and entomology in the eastern Mediterranean: Research into the history of insect synanthropy in Greece and Egypt*. Oxford: Archaeopress.

Panagiotakopulu, E., & Buckland, P. C. (1991). Insect pests of stored products from Late Bronze Age Santorini, Greece. *Journal of Stored Products Research, 27*(3), 179–184.

Payne, S. (1973). Kill-off patterns in sheep and goats: The mandibles of Aşvan Kale. *Anatolian Studies, 23*, 281–303.

Payne, S., & Munson, P. J. (1985). Ruby and how many squirrels: The destruction of bones by dogs. In N. R. J. Fieller, D. D. Gilbertson, & N. G. A. Ralph (Eds.), *Paleobiological investigations: Research design, methods, and data analysis* (pp. 31–39). (BAR international series 266)). Oxford: British Archaeological Reports.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpasian-Roodenberg, S., et al. (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One, 10*(6), e0129102.

Preece, R. G. (1981). The value of shell microsculpture as a guide to the identification of land Mollusca from quaternary deposits. *Journal of Conchology, 30*, 331–337.

Prehn, N., Feneru, F., & Rochester, J. (n.d.). *Beginner's guide to identifying British mammal bones*. London: Natural History Museum. https://www.nhm.ac.uk/content/dam/nhmwww/take-part/identify-nature/british-mammal-bones-ID-guide.pdf.

Prendergast, M. E., & Sawchuk, E. (2018). Boots on the ground in Africa's ancient DNA 'revolution': Archaeological perspectives on ethics and best practices. *Antiquity, 92*(363), 803–815.

Price, M., Wolfhagen, J., & Otárola-Castillo, E. (2016). Confidence intervals in the analysis of mortality and survivorship curves in zooarchaeology. *American Antiquity, 81*(1), 157–173.

Quitmyer, I., Hale, S., & Jones, D. S. (1985). Paleoseasonality determination based on incremental shell growth in the hard clam *Mercenaria mercenaria*, and its implications for the analysis of three Southeast Georgia coastal shell middens. *Southeastern Archaeology, 4*, 27–40.

Rackham, J. (1994). *Interpreting the past: Animal bones*. Los Angeles: University of California Press.

Rhodes, S. E., Ziegler, R., Starkovich, B. M., & Conard, N. J. (2018). Small mammal taxonomy, taphonomy, and the paleoenvironmental record during the Middle and Upper Paleolithic at Geissenklösterle Cave (Ach Valley, southwestern Germany). *Quaternary Science Reviews, 185*, 199–221.

Ringrose, T. J. (1993). Bone counts and statistics: A critique. *Journal of Archaeological Science, 20*(2), 121–157.

Rohland, N., & Hofreiter, M. (2007). Ancient DNA extraction from bones and teeth. *Nature Protocols, 2*(7), 1756–1762.

Russell, N. (2012). *Social zooarchaeology: Humans and animals in prehistory*. Cambridge: Cambridge University Press.

Sadighzadeh, Z., Tuset, V. M., Valinassab, T., Dadpour, M. R., & Lombarte, A. (2012). Comparison of different otolith shape descriptors and morphometrics for the identification of closely related species of Lutjanus spp. from the Persian Gulf. *Marine Biology Research, 8*, 802–814.

Salladié, P., Rodríguez-Hidalgo, A., Huguet Pàmies, R., Cáceres, I., Díez Fernández-Lomana, J. C., et al. (2014). The role of carnivores and their relationship to hominin settlements in the TD6-2 level from Gran Dolina (Sierra de Atapuerca, Spain). *Quaternary Science Reviews, 93*, 47–66.

Santana-Sagredo, F., Hubbe, M., & Uribe, M. (2016). Isotopic evidence for marine consumption and mobility in the Atacama Desert (Quillagua, northern Chile). *International Journal of Osteoarchaeology, 26*, 476–489.

Savelle, J. M., & Friesen, T. M. (1996). An odontocete (Cetacea) Meat Utility Index. *Journal of Archaeological Science, 23*, 713–721.

Sawada, J., Nara, T., Fukui, J., Dodo, Y., & Hirata, K. (2014). Histomorphological species identification of tiny bone fragments from a Paleolithic site in the northern Japanese archipelago. *Journal of Archaeological Science, 46*, 270–280.

Schiffer, M. B. (1976). *Behavioral archaeology*. New York: Academic.

Schiffer, M. B. (1987). *Formation processes of the archaeological record*. Albuquerque, NM: University of New Mexico Press.

Schmid, E. (1972). *Atlas of animal bones: For prehistorians, archaeologists, and quaternary geologists*. New York: Elsevier.

Serjeantson, D. (2009). *Birds* (Cambridge manuals in archaeology). Cambridge: Cambridge University Press.

Shapiro, B., & Hofreiter, M. (Eds.). (2012). *Ancient DNA: Methods and protocols*. New York: Springer.

Sherratt, A. G. (1981). Plough and pastoralism: Aspects of the secondary products revolution. In I. Hodder, G. Isaac, & N. Hammond (Eds.), *Pattern of the past* (pp. 261–306). Cambridge: Cambridge University Press.

Shufeldt, R. W. (1909). Osteology of birds. *Bulletin of the New York State Museum, 130*, 5–381.

Simonsen, K. P., Rasmussen, A. R., Mathisen, P., Petersen, H., & Borup, F. (2011). A fast preparation of skeletal materials using enzyme maceration. *Journal of Forensic Sciences, 56*(2), 480–484.

Smith, D. C. (Ed.). (1992). *Age determination and growth in fish and other aquatic animals*. East Melbourne: CSIRO Australia.

Spinage, C. A. (1972). Age estimation of zebra. *African Journal of Ecology, 10*(4), 273–277.

Spinage, C. A. (1973). A review of the age determination of mammals by means of teeth, with especial reference to Africa. *African Journal of Ecology, 11*(2), 165–187.

Steele, T. E. (2015). The contributions of animal bones from archaeological sites: The past and future of zooarchaeology. *Journal of Archaeological Science, 56*, 168–176.

Stewart, F. L. (1991). Floating for fauna: Some methodological considerations using the Keffer Site (AkGv-14) midden 57 faunal sample. *Canadian Journal of Archaeology, 15*, 97–115.

Stiner, M. C. (1994). *Honor among thieves: A zooarchaeological study of Neandertal ecology*. Princeton: Princeton University Press.

Stiner, M. C. (2002). On *in situ* attrition and vertebrate body part profiles. *Journal of Archaeological Science, 29*(9), 979–991.

Stiner, M. C., & Munro, N. D. (2002). Approaches to prehistoric diet breadth, demography, and prey ranking systems in time and space. *Journal of Archaeological Method and Theory, 9*(2), 181–224.

Stiner, M. C., Beaver, J., Munro, N. D., & Surovell, T. (2008). Modeling Paleolithic predator-prey dynamics and the effects of hunting pressure on prey 'choice'. In J. P. Bocquet-Appel (Ed.), *Recent advances in palaeodemography: Data, techniques, patterns* (pp. 143–178). Dordrecht: Springer.

Talmale, S. S. (2017). *Taxonomic studies of small mammals (Scandentia, Rodentia, Soricomorpha and Chiroptera) of Madhya Pradesh* (Records of the Zoological Survey of India Occasional Papers 383). Kolkata: Zoological Survey of India.

Talmale, S. S., & Pradhan, M. S. (2009). *Identification of some small mammal species through owl pellet analysis* (Records of the

Zoological Survey of India Occasional Papers 294). Kolkata: Zoological Survey of India.
Tchernov, E. (1968). *Succession of rodent faunas during the upper Pleistocene of Israel*. Hamburg: Mammalia Depicta.
Thomas, K. D. (1985). Land snail analysis in archaeology: Theory and practice. In N. R. J. Fieller, D. Gilbertson, & N. G. A. Ralph (Eds.), *Palaeobiological investigations: Research design, methods and data analysis* (pp. 131–156). (BAR international series 266)). Oxford: British Archaeological Reports.
Twiss, K. (2012). The archaeology of food and social diversity. *Journal of Archaeological Research, 20*(4), 357–395.
Vigne, J.-D., & Helmer, D. (2007). Was milk a "secondary product" in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. *Anthropozoologica, 42*(2), 9–40.
Vives, S., Gilbert, M. T., Arenas, C., Gigli, E., Lao, O., & Lalueza-Fox, C. (2008). Statistical analysis of post mortem DNA damage-derived miscoding lesions in Neandertal mitochondrial DNA. *BMC Research Notes, 1*(1), 40.
Walker, R. (1985). *A guide to post-cranial bones of east African animals*. Norwich: Hylochoerus Press.
Waselkov, G.A. (1987). Shellfish gathering and shell midden archaeology. Advances in archaeological method and theory, 10, 93–210.
West, C. F., Burchell, M., & Andrus, F. T. (2018). Molluscs and paleoenvironmental reconstruction in island and coastal settings: Variability, seasonality, and sampling. In C. M. Giovas & M. J. LeFebvre (Eds.), *Zooarchaeology in practice: Case studies in methodology and interpretation in archaeofaunal analysis* (pp. 191–208). New York: Springer.
Wilbur, K. M., & Yonge, C. M. (Eds.). (1964). *Physiology of Mullusca* (Vol. 1). New York: Academic.
Williams, S. L. (1999). *Destructive preservation: A review of the effect of standard preservation practices on the future use of natural history collections*. Waco, TX: CCI Printing Solutions.
Wolverton, S. (2013). Data quality in zooarchaeological faunal identification. *Journal of Archaeological Method and Theory, 20*, 381–396.
Wye, K. R. (1991). *The illustrated encyclopedia of shells*. London: Quantum Books.
Yeh, H. Y., Chen, Y. P., & Mitchell, P. D. (2016). Human intestinal parasites from the Wushantou site in Neolithic period Taiwan (800-1 BC). *The Journal of Island and Coastal Archaeology, 11*(3), 425–434.
Yom-Tov, Y., & Wool, D. (1997). Do the contents of barn owl pellets accurately represent the proportion of prey species in the field? *The Condor, 99*(4), 972–976.
Yonge, C. M., & Thompson, T. E. (1976). *Living marine molluscs*. London: Collins.
Zeder, M. A. (2012a). Pathways to animal domestication. In P. Gepts, T. R. Famula, & R. L. Bettinger (Eds.), *Biodiversity in agriculture: Domestication, evolution, and sustainability* (pp. 227–258). Cambridge: Cambridge University Press.
Zeder, M. A. (2012b). The domestication of animals. *Journal of Anthropological Research, 68*(2), 161–190.
Zeder, M. A. (2012c). The broad spectrum revolution at 40: Resource diversity, intensification, and an alternative to optimal foraging explanations. *Journal of Anthropological Archaeology, 31*(3), 241–264.
Zeder, M. A. (2015). Core questions in domestication research. *Proceedings of the National Academy of Sciences, 112*(11), 3191–3198.

16 Archaeological Plant Remains

At this period agriculture was not entirely unknown. This is proved . . . by the discovery of carbonised cereals at various points. Wheat is most common, having been discovered at Meilen, Moosseedorf, and Wangen. . . . many bushels of it were found, the grains being united in large thick lumps. In other cases the grains are free, and without chaff, . . . while more rarely they are still in the ear. . . . Still more unexpected was the discovery of bread, or rather cakes, for leaven does not appear to have been used. They were flat and round, . . . and, to judge from one specimen, had a diameter of four or five inches

(Lubbock 1865: 153–154).

As Lubbock's observations about waterlogged finds at prehistoric Swiss "lake villages" show, archaeologists have long used remains of plants and plant products as evidence for aspects of ancient societies. Archaeobotany, also known as paleoethnobotany, is the subdiscipline of archaeology concerned with the contributions of plant remains to our understanding of past behaviors, diets, food preparation and culinary practices, environments, and technologies. Both terms differ from paleobotany, which also involves the study of ancient plant remains, but without explicit connection to human history. There are also terms for specialized aspects of archaeobotany, such as archaeoanthracology (study of archaeological charcoals) and archaeological palynology (study of archaeological pollens). Archaeobotany is very collaborative in that a holistic understanding of the botanical traces benefits from allied information from zooarchaeologists, paleoecologists, pottery and lithic specialists, and others.

16.1 Types of Archaeological Plant Remains

Archaeobotanists typically classify plant remains as plant macroremains, microremains, and chemical and isotopic traces, each with distinct roles in our understanding of the past as well as distinct challenges with respect to preservation, data collection, quantification, and interpretation (Madella et al. 2014; Pearsall 2015, 2019).

16.1.1 Macroremains

Macroremains are visible to the naked eye and their identification typically requires no more than low-power magnification. Among this type of evidence are charcoal, nuts and nutshell, tubers (parenchymous organs), fruit endocarps (seeds), or their fragments, plant parts closely associated with seeds (e.g., glumes and awns of cereals), and pieces of wooden artifacts or basketry, each with different probabilities of preservation in various environments.

16.1.2 Microremains

Plant remains that are visible to the naked eye are often less abundant than microscopic pollen, spores, opaline phytoliths, diatom frustules, and starch grains. These microremains also offer different kinds of information than macroremains and have different taphonomic histories. Consequently, they are a good complement to plant macroremains and other kinds of evidence.

Pollen grains are an essential element in the reproduction of flowering plants. Male gametes (sperm cells) occur inside each pollen grain or pollen tube, and the pollen transmits male genetic material to a female gamete during pollination by gravity, wind, water, insects, or human agency. The wall, or *exine*, of the pollen grain contains a substance, called sporopollenin, that is very resistant to decay. This makes it

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_16

able to survive for long periods in some cases, providing a very useful record of the past abundances of flowering plants.

Phytoliths result when some plants deposit silica (and sometimes calcium oxalate) that they have taken up from the soil onto cell boundaries. Phytoliths vary considerably in shape, even among different plant parts, and inter-specific differences in cell structure may make phytoliths a good source of evidence for plant taxa (Fig. 16.15). Because the silica survives long after other parts of a plant's anatomy have decayed away, phytoliths can provide archaeological evidence for the use of plants or plant parts that are not represented among charred macroremains. Phytoliths can occur in sediments, but also in dental calculus and on the surfaces of tools and pottery. In the case of calculus on human teeth, they are providing direct evidence of what plants an individual has chewed.

Starch is the material in which plants store carbohydrate for later use to make energy, which is what makes it a potential food source for animals as well. Its granules are composed of two kinds of polymer molecules, called amylose and amylopectin, which occur in alternating crystalline and amorphous layers in the grains. Starch granules can be preserved on the surfaces of pottery or stone tools or that came into contact with plants, and in dental calculus.

16.1.3 Chemical and Isotopic Evidence

Chemical and isotopic evidence for prehistoric plants includes residues of fatty acids and amino acids on surfaces or edges of artifacts or on the surface or in the matrix of pottery (pp. 204; Evershed et al. 1992). It also includes DNA fragments in plant remains, and carbon isotopes and trace elements in bones whose relative abundances in part reflect the plants in an animal's or person's diet.

16.2 Taphonomy, Site-Formation Processes, and the *Chaînes Opératoires* of Plant Use

As in zooarchaeology (pp. 242–243), archaeobotanists must concern themselves with the taxonomy of the organisms they study and with the taphonomy of their remains (Beck 1989; Gallagher 2014; Hubbard and al-Azm 1990; Pearsall 2019). Identification of plant remains to genus or species faces challenges because usually only tiny remnants of ancient plants survive in archaeological deposits, where they are preserved at all. By contrast, botanists studying modern plants usually have whole plants at their disposal, and taxonomists have tended to emphasize the characteristics of leaves and flowers, parts that archaeologists are unlikely to find. This often requires archaeobotanists and paleobotanists to create their own studies of the morphology and structure of the plant parts that survive more regularly, and in the state in which they are typically found. Dry caves and waterlogged sediments are generally favorable for preservation of many kinds of macroremains, because they exclude or impede bacterial decay. Charring in a reducing atmosphere, which may caramelize the sugars in plant materials so they are unattractive to bacteria, enhances preservation in other kinds of sediments. Macroremains may also be preserved in ancient feces (coprolites, e.g., Minnis 1989), as impressions in pottery or bricks or, rarely, as casts in volcanic ash (Farahini et al. 2017).

Much as with tool-making, it is possible to conceive of plant use in a *chaîne opératoire* that includes acquisition, not only of plants through harvest in the wild or in fields, but also of the tools and other materials necessary for their collection, processing, and use. For cultivated plants, the *chaîne opératoire* could even include decisions, activities and technologies related to sowing, planting, and cultivating plants, when and where to sow them, fallow schedules, and which crops to grow together. Just as with lithics, the decisions in this *chaîne opératoire* yield products and by-products that provide clues to processes involved up to and including the plants' use and discard. The *chaîne opératoire* thus influences plant remains' taphonomy.

Hillman (1984) pioneered the idea that the sample assemblages of plant remains we are able to recover archaeologically have an intimate connection, not only to the post-depositional processes that gradually alter deposited assemblages, but also to the harvesting, processing, use and discard decisions and activities that resulted in those deposited assemblages. Focussing mainly on cereals and similar crop plants, such as peas and vetch, he used experiments and ethnographic observations in Turkey to characterize the numerous steps from harvest to use and discard, the most likely products of each step, and the relative probabilities of preservation of each product (e.g., Fig. 16.1).

In this methodology, the "life assemblage" (see p. 242) of plants, or even the absolute amount of plant food consumed, is not typically of much interest and is impossible to reconstruct from the number of macro- or microremains in a sample (Pearsall 2019: 61). Rather, archaeobotanists are more interested in various kinds of deposited assemblages that represent material lost or discarded at different stages of some *chaîne opératoire*. For example, in processing glumed cereals, one possible product is straw waste (Table 16.1). If the straw store is accidentally burned or is used to temper dung fuel, this could result in a deposited assemblage of charred culm nodes, awn segments, basal spikelets, weed heads larger than the grain spikelets, and weed seeds smaller or the same size as grain. Alternatively, there could be plant impressions of these items if the straw was used as temper in pottery or bricks (see also Smith 2001).

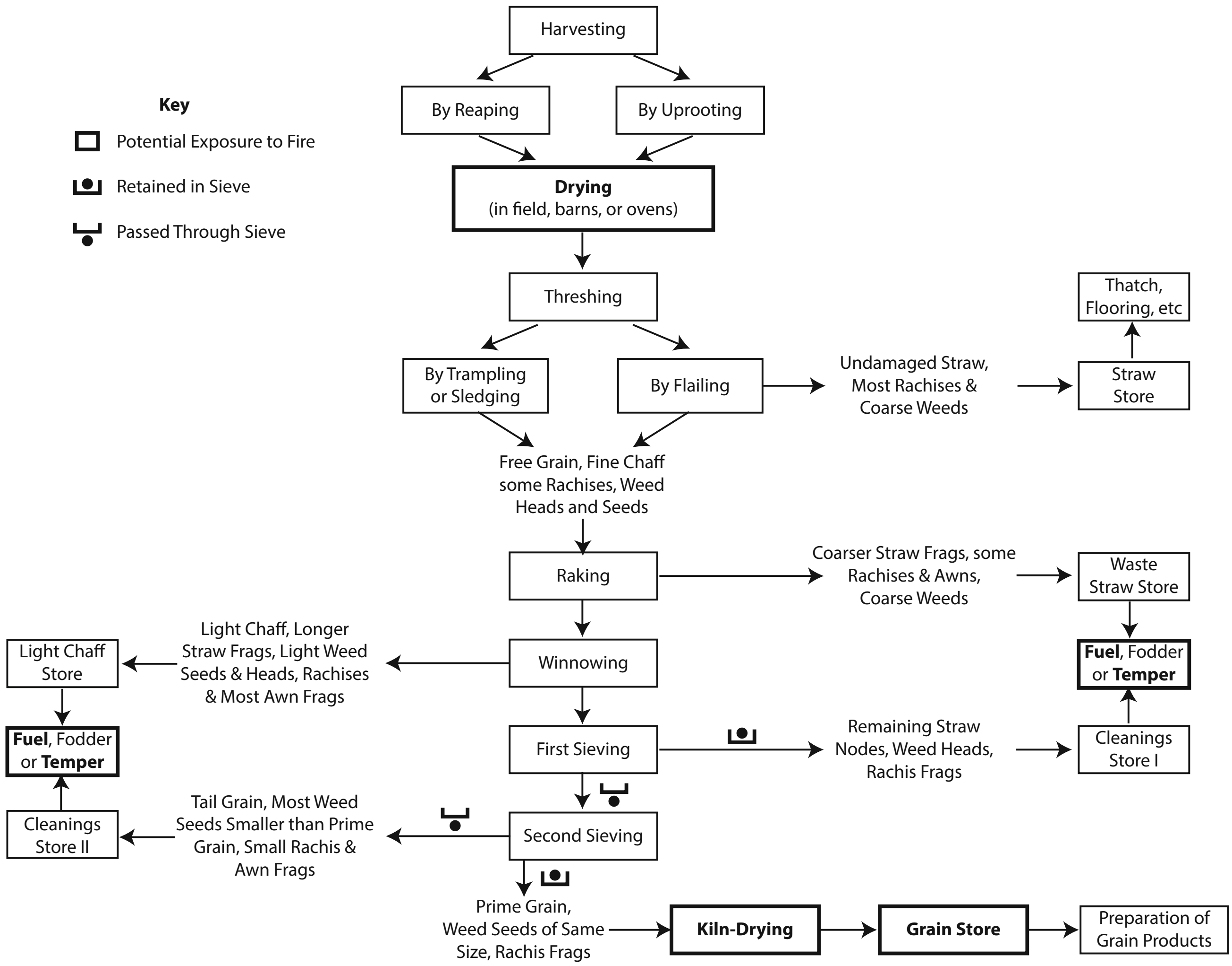


Fig. 16.1 A simplified flow chart of processes involved in the harvest and processing of free-threshing cereals and their straw in traditional agriculture in Turkey. (Modified from Hillman 1984: 4). Note that the products or by-products that result from some processing steps have some probability of accidental preservation through charring, while others have very little probability of surviving at all

Consequently, Hillman relies, not on counts of plant remains as some indirect measure of the number of plants in a life assemblage or harvested assemblage, but rather on ordinal-scale measures on a variety of plant parts that, in combination, provide clues to the plant-processing activities that probably created the deposited assemblage. These include the seeds of weeds whose sizes would lead to their removal by sieving and others whose sizes are so close to those of crop seeds that they could only be removed by hand-picking.

The *chaînes opératoires* of processing other kinds of plants are quite different from those of cereals in the example taken from Hillman's work, and have different taphonomic outcomes. In addition, the deposited assemblages most likely to preserve are different when we look for evidence other than charred macroremains. The conditions that favor phytolith preservation, for example, are imperfectly understood but are not enhanced by fire, and preservation varies among phytolith types (Cabanes and Shahack-Gross 2015; Cabanes et al. 2011; Piperno 2006: 21–22).

However, for some kinds of plant evidence, especially from pollen and charcoal, it is often the life assemblage that is of interest. The relative abundances of plant taxa as represented by offsite pollen from cores in lakes and bogs and charcoal from archaeological sites can tell us something about the populations of trees and other flowering plants in the vicinity of the site and the much larger region that surrounds it. They thus provide indirect evidence for climate change and forest clearance. However, useful interpretation of pollen abundances requires careful consideration of the processes that transported pollen to the sampling locations and variation in the pollen productivity of the plants, as well as their probabilities of preservation. In the case of charcoal from archaeological sites, we need to consider differential combustion, probable selectivity by humans in their

Table 16.1 Simplified composition of expected contents of crop product types from glume wheats (after Hillman 1984: Table 1). Ordinal scale: items (e.g., glume and rachis fragments) result from the fragmentation of larger items (e.g., spikelets) after charring has occurred

Crop product type		Culm bases	Culm nodes	Awn segments	Intact non-basal spikelets	Intact basal spikelets	Spikelet forks	Glume bases
ΣB	Residue from burning whole sheaves	xx	XX	XX	XXX	XX	XXX	XXX
B1	Straw Waste from raking, winnowing and coarse riddling (threshing floor waste)	xx	XX	XX	r	XX		
B2+B4+B5	Threshed spikelets charred during parching or accidental burning of spikelet store	x	X	X	XXX		XXX	XXX
B2	Coarse sievings (larger than prime grain)	x		X	XX			
B3	Fine sievings (smaller than prime grain)	r	r				XX	XXX
B4	Semi-clean grain in bulk storage charred by sterilization or accidental burning	r	r	r			XX	XXX
B5	Cleanings from hand-sorting before food preparation	r	r	r			X	X
B6	Clean prime grain charred during roasting							

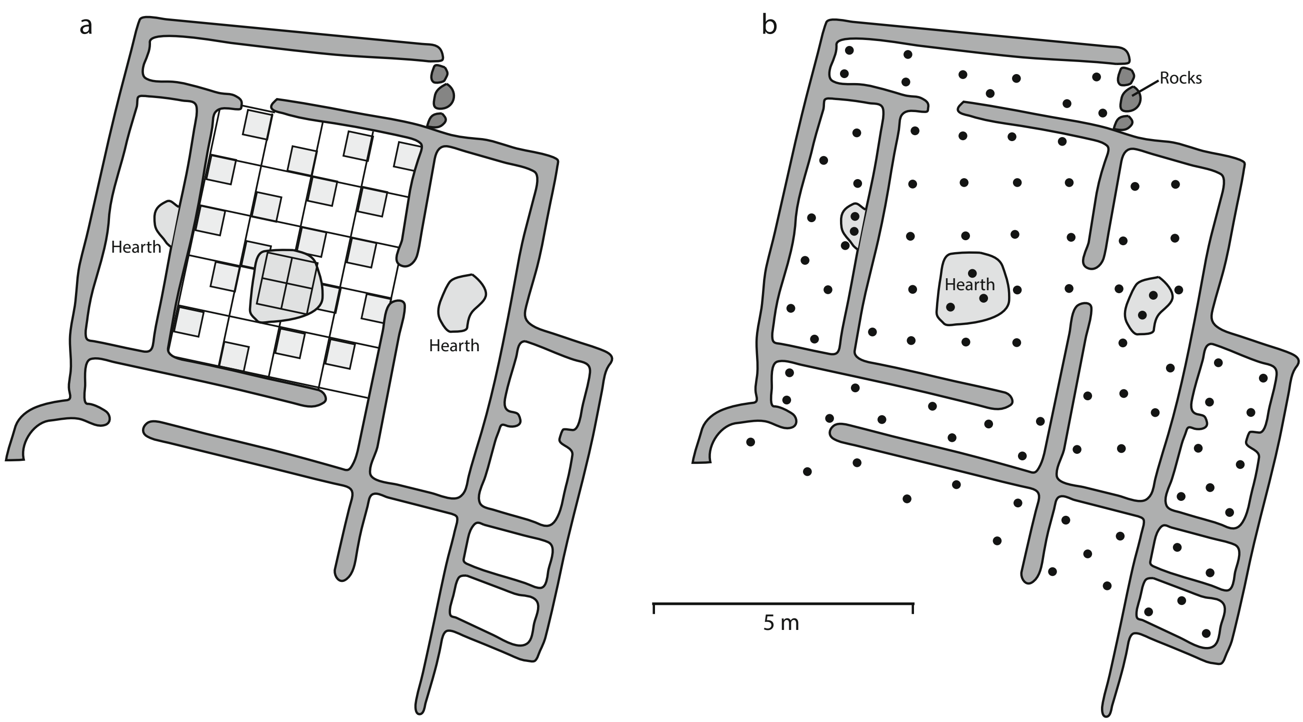


Fig. 16.2 Bulk sampling (**a**) of an excavation area's deposits and features as compared with pinch sampling (**b**; plans after Metcalfe and Heath 1990). The bulk sample in (**a**) is a systematic, stratified, unaligned cluster sample of a population consisting of the floor in one room, with a hearth sampled separately or treated as a stratum in a stratified sample, and sample size of $n = 19$ for the floor or $n = 23$ with the hearth included (see Chap. 6). The pinch sample is also a cluster sample, but in this case each room floor or hearth is a population and we combine the pinches from dispersed locations (filled circles) to gather sufficient sediment from each room or hearth to ensure adequate numbers of plant remains for statistical purposes. In that case, the sample size for each room or hearth is only $n = 1$. Alternatively, if we treat the whole building as the population, the sample size is $n = 14$ (3 hearths and 11 room floors)

XXX = abundant, XX = common, X = few, r = rare, lower case x means it only occurs when harvest was by uprooting (culm bases). Note that many

Non-basal glume frags	Rachis internode frags	Prime grain	Tail grain	Weed heads > Spikelets	Weed heads ~ Spikelets	Weed heads < Spikelets	Seeds > Prime grain	Seeds ~ Prime grain	Seeds < Prime grain	Most common groupings
XXX	XXX	XXX	X	XX	XX	X	X	XX	XXX	
				XX			X	XX	XXX	
XXX	XXX	XXX	X		XX	X	X	XX	XXX	
						X	r			
	XXX		XX						XXX	
	r	XXX	X					XXX	X	
	r							XXX	X	
		XXX	X							

collection of wood for fuel or building material, which could include not only wood from nearby forests, but potentially wood imported from elsewhere, natural processes that transport charcoal from offsite or between stratigraphic layers, and archaeologists' own sampling methods (Höhn and Neumann 2018; Jansen and Nelle 2014; Marguerie 2002; Smart and Hoffman 1988: 168–170).

16.3 Archaeobotanical Sampling

As discussed more thoroughly in Chap. 6, ensuring that the sample we use to make inferences about sites, assemblages, features or other "populations" is representative requires careful attention to research design. In archaeobotany, whether the research questions involve determining when certain plants became domesticated in a site or region, inferring the diets or land-use practices of a site's inhabitants, investigating the cultural impacts of climate change, or identifying the processing stage associated with the plant remains in a deposit has a necessary relation to the sampling strategy, which begins in the field but continues in the laboratory. A major consideration is whether investigation of the research problem involves a spatial component (e.g., patterning across a site or differences among houses or features), changes over time, or characterization of whole sites, stratigraphic phases within sites, or the landscapes around sites. Anticipated taphonomic factors and the substantial labor requirements of processing, sorting, and identifying plant remains also have important impacts on sampling strategies.

Not surprisingly, archaeobotanists have given considerable attention to the merits and disadvantages of different kinds of samples. Much of this discussion has focused on the differences between "bulk" and "pinch" samples, usually divorced from an explicitly statistical perspective.

Many archaeobotanists recommend what they call "blanket sampling," meaning that there should be at least one sample element from every context—layer, floor, or feature—in every excavation unit that was excavated. Statistically, this could be a stratified sample (p. 97), each type of context corresponding to a stratum, and, for most excavations, there would be many strata and disproportionate sampling (i.e. some strata would have larger sampling fractions than others). Each sample element consists of a volume of sediment, and it is generally recommended, as far as possible, to make this a consistent volume, such as 10 L, but it is in any case essential to record what the volume was. One reason for favoring blanket sampling is that, if you sample only contexts where you predict there will be lots of plant remains, such as pits or hearths, your prediction could well be incorrect. Furthermore, estimates based on this sample could be biased, as the pits or hearths may not be representative of all the plant remains in the population of interest. At a minimum, use of a more restrictive sampling strategy should only occur with the recognition of what the population is (the population of hearths, or population of pits). Whether fieldworkers employed blanket sampling or not, there are multiple ways to obtain these sample elements (Pearsall 2019: 41–48).

Bulk sampling, including a version that is sometimes called point sampling, is a form of cluster sampling (see

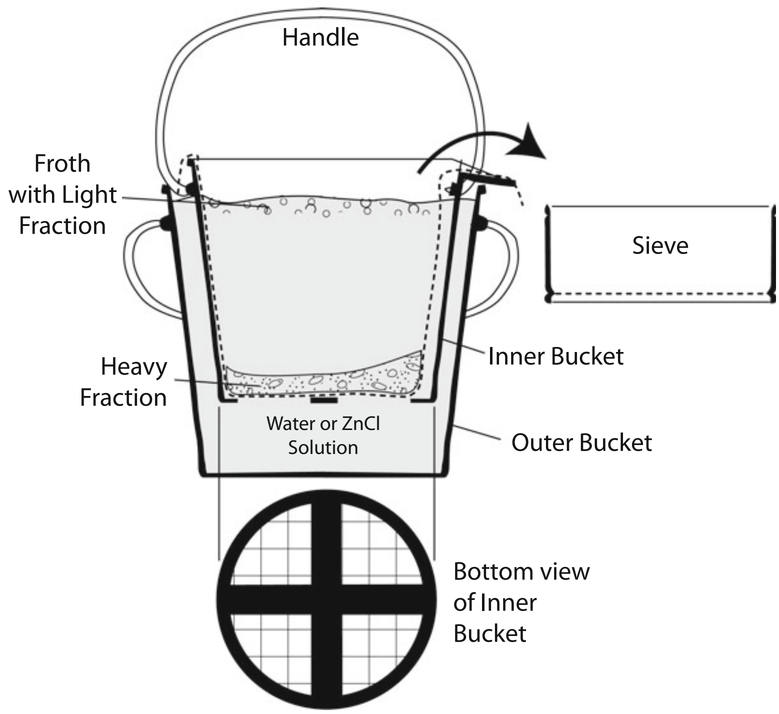


Fig. 16.3 Schematic of a simple two-bucket flotation system. The upper bucket is pierced in four quadrants to allow water to pass through the bottom mesh and pumping the upper bucket up and down with alternating rotation helps to churn up the sediment and create froth. The upper bucket may have a spout to facilitate decanting the froth into an adjacent screen, as here, or the user may simply skim light fraction from the froth with a small sieve

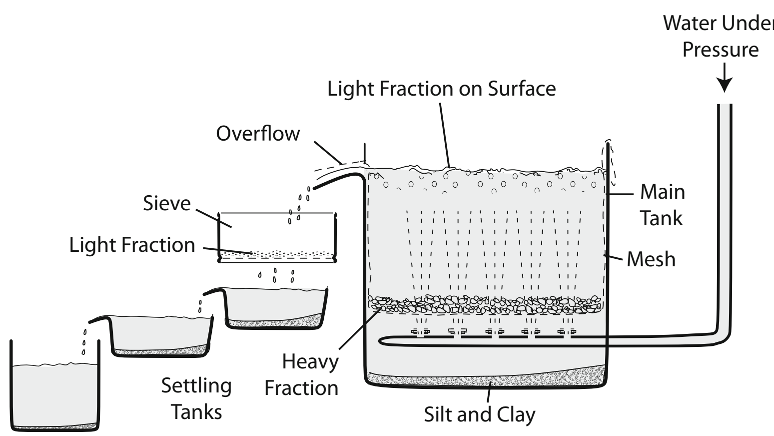


Fig. 16.4 Schematic of a simple froth flotation system with gravity pressure to create underwater jets to agitate the tank, and a series of settling basins to permit recycling of water. (E. Banning and Y. Salama)

pp. 98–100) with sample elements that consist of continuous volumes of sediment, such as all the sediment in a 1 cm layer across a square in a grid, or all the sediment in the bottom of a silo. The selection of volumes for bulk sampling can be purposive, such as targeting features like hearths, pits and silos, or based on probability sampling, such as stratified random sampling with strata that consist of different kinds of deposits and features, although this is not consistent with blanket sampling. Other alternatives are systematic sampling or systematic, stratified, unaligned sampling of a grid across a room's floor or campsite's or midden's surface (Fig. 16.2a). The statistical assumption of cluster sampling is that each bulk volume is something like a well-mixed microcosm of the population or spatial location it represents and, when the sampling fraction is large, as when the sample element contains most of the sediment that can be securely associated with a 50 cm × 50 cm grid unit, that assumption is probably reasonable. When it is a spatial sample, such as a systematic sample on a grid, the volume is sometimes taken as representative of a point at the grid unit's center. It is essential to measure the volumes of sediment and best to keep them reasonably consistent in size, but these volumes can be rather small (e.g., 0.6–2 L). Where the individual volumes are used as sample elements for a larger population, such as an entire room or layer in a site, they provide a basis for estimating things like ubiquities, densities, and standard errors.

Composite, scatter, or **pinch sampling** involves attempts to ensure that sample elements are representative by collecting small volumes of sediment from different parts of a context but then mixing them together in the same bag. This might be appropriate in cases where, for example, you want to represent the plant remains over a larger area, like a room floor, but do not have the time or resources to divide it into a grid for bulk or point sampling or to analyze the multiple sample elements that would result from that (Fig. 16.2b). Where possible, fieldworkers should collect enough pinches to fill a standard volume, such as 10 L. We could statistically characterize pinch sampling as either a kind of subsampling or a case of the context—the room floor—being the population and the "pinches" being the sample elements, except that we then mix them to make a single sample element. This may result in a more representative sample of sediment from that context than would a bulk element of equal volume from a single location in the context. However, physically mixing the pinches together prevents us from identifying differences between the pinches that are essential for calculating standard deviations and standard errors. Many archaeobotanists like the ease and time savings of this method, often prefer the larger volumes (~10 L) that result, and are willing to sacrifice the ability to calculate confidence intervals for their data, except in cases where the population consists of multiple contexts sampled in this way. In these last cases, the sample size is the number of contexts or 10 L bags, and *not* the number of pinches.

Column sampling is used in cases where the goal is to identify changes in some characteristics of the plant remains over time. It involves either randomly or, more often, purposively selecting locations in which to take a sequence of sample elements from bottom to top of a series of sediments. Typically, a fieldworker selects a location along a "baulk," a vertical section through the stratification in an excavation area that includes several superimposed stratigraphic layers (see Chap. 19), and takes a small volume of sediment from each layer with a trowel or spoon, taking care to avoid the boundaries between layers (in contrast to micromorphology sampling, see pp. 295, 302). In such cases, it is important to sample from bottom to top to prevent contamination of lower sample elements with particles falling from above. Another version of column sampling is to take successive sample elements by augering, usually in arbitrary vertical intervals, like 10 cm, or to take sections from cores (see pp. 294–295). Disadvantages of column sampling include that the volumes of sample elements tend to be very small (~ 0.1 L) and thus unlikely to contain very many macroremains (although they may contain many microremains), and that one column or a small number of columns may not provide a very representative sample of a target population that could be a very extensive layer in a site.

Laboratory Sampling While the most significant sampling decisions may occur in the field, often the sheer volume of material recovered in the field is so large that archaeobotanists need to subsample (multistage sampling). Generally, it is a good idea in these instances to take a stratified random sample that ensures that each category of context contributes to the estimates of population parameters (p. 97). Alternatively, it may be advisable to take a sample or samples that contribute to evaluating very specific research questions, such as whether there was a change in selection of fuels (two or more populations of hearths) or a shift from dry-farming to irrigation-farming (populations of different ages). For charcoal, it is also important to ensure that the sub-sample represents a range of fragment sizes, since charcoal of some taxa may tend to occur in smaller fragments than others (Smart and Hoffman 1988: 174–176). However, there could still be some selection bias, as larger fragments are often easier to identify to a higher taxonomic level.

Sample Size Archaeobotanists have only rarely used statistical methods to determine fixed sample sizes (but see Pearsall 2015: 106–1078; van der Veen and Fieller 1982; and see Case Study: Plant Remains from Sites in Korea) but have often used a form of sequential sampling (p. 101) to provide samples that are adequate for their purposes. This is usually "sampling to redundancy," which means gradually increasing sample size until the sample's richness— i.e., the cumulative number of taxa—levels off. What this means is that further increases to sample size are unlikely to yield any taxa not already included in the sample. A similar approach is to increase sample size until the proportions of the main taxa level off, indicating that addition of further sample elements is unlikely to change those proportions substantially (see Fig. 6.6).

In selecting an archaeobotanical sampling method, including both sample size and the size of individual sample elements (see pp. 100–102), it is important to anticipate the likely density of the seeds or other plant materials that are important to the research questions. For quantitative analyses, the sample elements should be large enough that the average counts of at least the more important taxa are not too close to zero, and you can model the expected counts with Poisson distributions (pp. 132–133) to determine whether 0.5 L, 1 L, 5 L or 10 L sample elements will be sufficient for the macroremains or microremains of interest. Alternatively, a period of trial and error may help to decide appropriate volumes. For qualitative analyses that are based on ubiquity measures (see pp. 122–124) or just the simple presence of certain taxa, density is also the most important determinant of the sample's usefulness. On the basis of estimates of seed densities, for example, you can use the binomial model (pp. 131–132) to decide on the optimum combination of sample size and sample element to provide reasonable assessments of whether a particular taxon is present or not, or the value of its ubiquity.

16.4 Processing Samples of Macroremains

Today, the usual way to retrieve plant macroremains from archaeological sediments is by water separation through a technique called flotation. Captured remains are then sorted and analyzed in the laboratory. Notably, not all of these remains will be plant remains; small bone fragments, snail shells, or stone flakes need to be passed on to the relevant specialists.

Flotation separates charred materials from the mainly inorganic sediment matrix because most charred remains are less dense than water, causing them to float. In combination with agitation, this concentrates them in the "light fraction" for easier analysis. Some heavier charred material, especially fruit stones and larger pieces of charcoal, sink into the "heavy fraction," which may also contain small artifacts, bone fragments, shells, and stones. Both fractions then require manual sorting to pick out identifiable plant remains.

16.4.1 Water Flotation for Charred Plant Remains

By taking advantage of the lower density of organic as compared to inorganic particles, flotation in water separates charred plant remains from the rest of sediment more effectively than dry sieving in most situations. Some analysts add sodium bicarbonate to the water to improve the removal of clay, calcium carbonate and other minerals that may adhere to the surfaces of plant remains. It may also recover some very small animal bones and lithics.

The simplest kind of flotation—manual **bucket flotation**—is good for processing units of sediment less than about 5 L in volume and, by exposing the plant remains to less prolonged and vigorous wetting, is less destructive than most alternatives (Fig. 16.3). It requires no equipment more sophisticated than several sizes of geological sieves, a small tea-strainer type of sieve, and one or two buckets (Fig. 16.4). In its simplest form, bucket flotation involves putting about a liter sediment into a bucket of water, swirling it around and then decanting the muddy water into the nested sieves. A slightly more sophisticated version involves two buckets, one of which has had most of its bottom cut out and replaced with a mesh. With the mesh-bottomed bucket nested into the intact one, the user fills them about two-thirds full of water and empties a volume of sediment, typically a liter, into the upper bucket. Grabbing the handle of the mesh-bottomed bucket, the analyst pulls it up, plunges it down, and rotates it back-and-forth to agitate the water, put the sediment into suspension, and release lighter particles from the sediment matrix so that they float while silts and clays sink and pass through the mesh at the bottom, while large, heavy particles, potentially including nutshell or lithics, are caught on the mesh. One either collects floating material with a small sieve or decants the muddy water into nested sieves and then extracts the heavy fraction from the mesh at the bottom before placing the captured material into bags or bundles for drying. Failure to dry samples thoroughly before transporting or analyzing them can result in severe breakage, crushing, or attack by fungi and mildew.

Bucket flotation is tedious for analysis of large volumes of sediment and may have very low yields of plant remains where their density is low. Consequently, many archaeobotanists use flotation machines in which air bubbles or water jets agitate the sediment in water (Hosch and Zibulski 2003; Pearsall 2015, 2019; VanDerwarker et al. 2016). **Machine-assisted flotation** facilitates processing of very large sediment volumes, such as the 10 L that some archaeobotanists favor.

For example, the Ankara-style flotation tank (Fig. 16.4; cf. Shell Mound Archaeological Project, or SMAP-type, flotation, Watson 1976) agitates sediment with jets of pressurized water—gravity is sufficient to provide the pressure through having the source water at a higher elevation than the tank—forced through small holes in interior pipes to agitate water in the tank. A mesh submerged in the tank captures the heavy fraction, and a lip or spout allows overflowing water and any floating material to exit into a series of nested sieves. In regions where water is scarce, the overflow water can pass through a series of settling tanks and be recaptured for reuse. After pouring a sediment sample of known volume and mass into the tank and stirring, the analyst simply collects floating material, most of it organic, on the sieves and then removes the internal mesh bag containing all of the heavier particles too large to pass through the mesh. Flotation thus removes all the finest particles and separates the lightest particles that remain from the rest. The captured light and heavy fractions are then put in separate small gauze or cloth bags or bundles and hung where they will dry slowly.

Flotation is not always the best method to use, as agitation in water may destroy some kinds of delicate bones and plant remains. Users of flotation tanks must also be careful to avoid contamination from previous samples in the tank. In some cases, dry sieving may be preferable.

16.4.2 Dry Sieving

Some archaeobotanists report that dry sieving is less likely than flotation or wet-screening to fragment fragile charred remains, although it may have lower recovery of very small seeds (Chiou et al. 2013). As with microrefuse and sediment analyses (Chap. 17), sieving typically involves use of nested screens, with the largest apertures at the top and smallest at the bottom, to sort sediments by particle size. Often a mechanical screen-shaker helps to make the sediments move downward through these screens, potentially damaging some fragile plant remains and generally creating a lot of dust, so that use of dust masks is a sensible precaution. Archaeobotanists often focus on the screen sizes that are most likely to catch charred seeds or husks of the plants of greatest interest.

16.4.3 Post-separation Analysis

Once plant macroremains have been separated from sediment, a low-power (7x–30x) binocular microscope is used to examine each fraction in small portions on a Petrie dish, where the analyst uses tweezers or a fine brush to sort particles into categories, such as charcoal, nuts, seeds, fruit stones, and unidentifiable or uncertain plant material. There can also be considerable non-plant material, such as lithics, sherds, and stones in the heavy fraction, and shells or insect parts in the light fraction. Sorting can take many hours per kilogram of charred material. Further sorting of each major category to genus or species can be even more time-consuming. Consequently, characterization of macroremains sometimes calls for subsampling, especially where the volumes of sediment sampled are very large.

Identification of macroremains requires a voucher collection with examples of wood, charred wood, seeds, nuts, and other plant items that you might expect to find in the archaeological deposits (Bye 1986; Pearsall 2015). Most archaeobotanists build their own collections, as the specimens in herbarium collections do not typically emphasize the parts that are of greatest archaeological interest, and their curators are understandably protective of specimens. Some kinds of voucher specimens, such as maize kernels, wheat grains, peas and lentils, are easy to obtain, at least in modern varieties, from bulk food stores, while it may be possible to obtain others from carefully identified plants in the wild or in farmers' fields. Wood of the more commercially valuable species may be obtainable from scrap at a lumberyard or carpentry shop, but collecting dead limbs of living trees can often yield a collection more similar to archaeological specimens. As charring changes metric attributes of seeds through shrinkage (Hubbard and al-Azm 1990), it is helpful to char some of the material from each taxon and plant part in a hearth or lab oven, preferably with an atmosphere that excludes oxygen, to make voucher specimens more comparable to archaeological ones.

16.5 Seeds and Nutshell

The remains of fruit, most often charred or desiccated seeds, nuts or nutshell, or their impressions in clay, have long been the most important type of plant macroremain, both because seeds and nuts have been important human food sources, and because they are more likely to survive archaeologically than many other kinds of plant parts. Plant parts adjacent to seeds, such as glumes and rachis fragments of cereals (Fig. 16.7), may also be preserved. Characteristics of some seeds and nuts or their uses make them good candidates for charring, and thus for enhanced preservation. For example, hulled wheats that need to be parched with heat before threshing are fairly likely to be charred (Harlan 1967). In other cases, preservation depends on intentional or accidental firing, either after discard in a hearth or when a building burned down. Only plant parts that were protected from direct burning are likely to be preserved, with enough heat to char them but not enough to combust them and reduce them to ash (Dimbleby 1967). For example, relatively dense nutshell discarded in fires tends to sink in ash, where it is protected from combustion.

Fruit and seeds of flowering plants are formed from the flowering parts of plants after fertilization of the egg cell (in the ovule) by pollen (Figs. 16.5, 16.6, 16.7, 16.8, and 16.9). As seeds grow, the ovary becomes enlarged, the flower's stamens and petals shrivel and fall off, and the ovary becomes recognizable as a fruit. Fruits vary widely in form, ranging from pods and dry capsules to fleshy, edible fruits, such as apples, tomatoes, cucumbers and bananas. Blackberries and raspberries are actually clusters of many

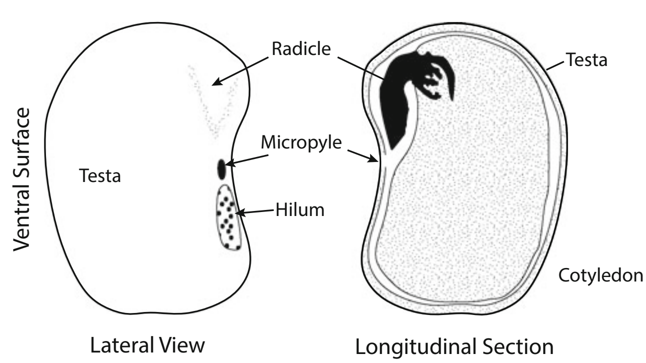


Fig. 16.5 Anatomy of a dicotyledon (dicot) seed. The embryo consists of a root or radicle and a shoot, or plumule, and is attached to two cotyledons. A cotyledon is a modified leaf in the seed, and monocots have one cotyledon, while dicots have two. The cotyledons store food and enclose the embryo. The hilum is a scar marking the place where the seed was attached to its pod

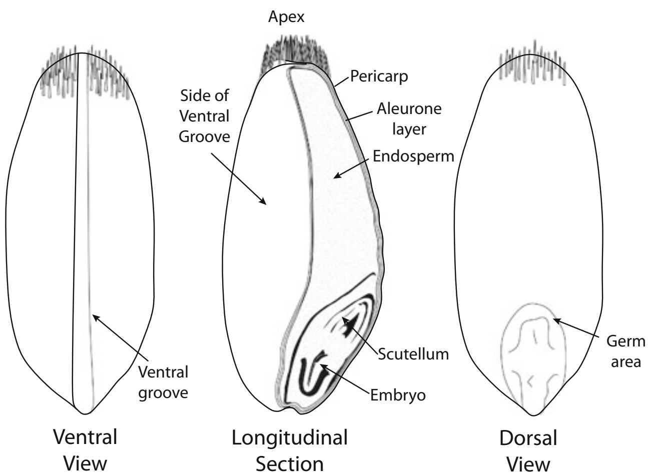


Fig. 16.6 Ventral and dorsal views and longitudinal section through a wheat grain (or caryopsis). As monocots, wheat plants store energy in a tissue called endosperm. (Y. Salama)

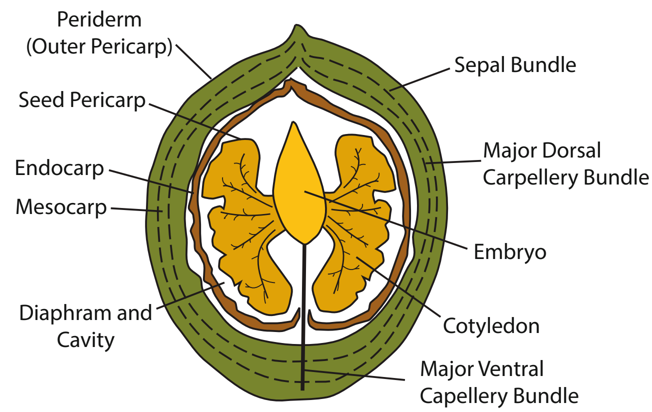
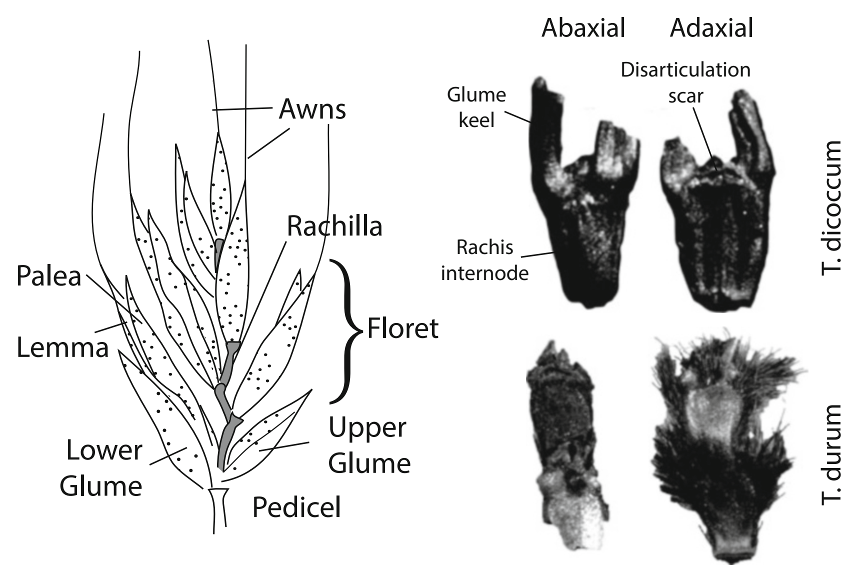


Fig. 16.9 Radial section through a walnut (*Juglans regia*) to show its parts

Fig. 16.7 Anatomy of a cereal spikelet with glumes enclosing florets attached by rachillae (left). The photos at right show both abaxial (away from the axis) and adaxial (toward the axis) views of charred rachis nodes and internode segments of *Triticum dicoccum* (top) and *Triticum durum* (bottom) from Bronze Age sites in the Near East. The *T. dicoccum* is a spikelet fork, as the bases of two glumes are present. The preservation of the *T. durum* at lower right is quite exceptional, with even very fragile hairs present. (Photos courtesy J. McCorriston)



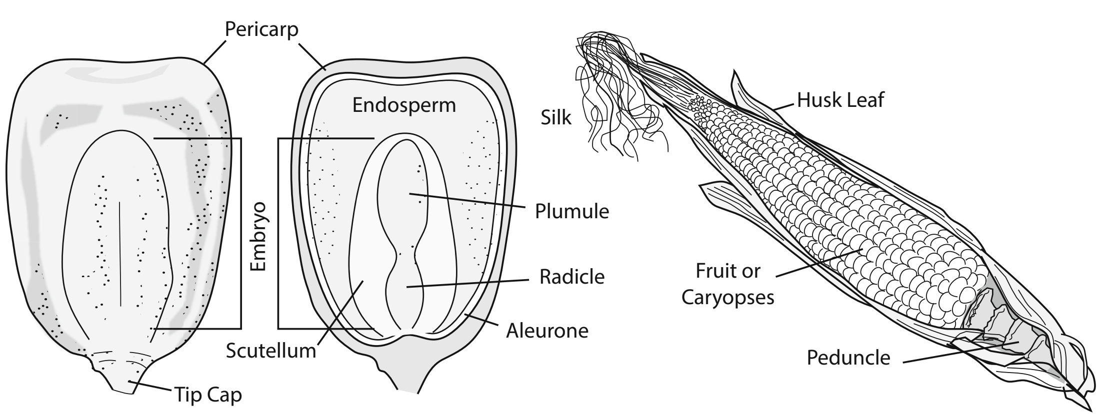


Fig. 16.8 Dorsal view and cross-section of a maize (*Zea mays*) caryopsis (kernel) and, at right, ear with female inflorescence. The endosperm stores energy for the embryo in the form of starch

small fruits. In strawberries, the largest edible portion is actually the flower's receptacle while the fruits are the small pips that cover it.

Sometimes, we are fortunate in finding other parts of a grass plant. Grains grow at the top of a culm (stem) in a spike (e.g., wheat and barley), broad panicle (e.g., oats), or intercalary ear (maize). A spike consists of a central, segmented rachis, with spikelets attached alternately at the nodes. A panicle is the same except that branches at each node are attached to the spikelets. Each spikelet (Fig. 16.7) has two glumes enclosing several florets attached by rachillae. In each floret, a lemma and palea enclose the grain or caryopsis, the lemma on the dorsal side and the palea on the ventral.

Maize (Fig. 16.8) is a highly unusual grass, being the product of humans' domestication of teosinte (Huffard et al. 2012). Here, the axis is a tough cob to which the grains are firmly attached. The glumes, lemma, and palea have become so reduced in size that the grains or kernels are naked and, instead, the entire corn ear is enclosed by leaf bases (husks). These completely preclude natural dispersal of the grains, making the plant completely dependent on humans for reproduction.

Nuts were exploited not only for their oil and meat, which can be made into flour, but also for tannic acid (for tanning hides) and fuel. Use of nutshell for fuel as well as roasting to reduce nuts' tannic acid content accounts for their charring in many archaeological cases. The part of the nut that is most often preserved archaeologically is the shell or **endocarp**, which in many kinds of fruit, such as peaches and walnuts, encloses the seed (endosperm and embryo) and lies beneath the fleshy part or mesocarp (Fig. 16.9).

Some guides to the identification of seeds and nuts include Delorit (1970); Nesbitt (2009), Schopmeyer (1974), and Smith (2018), but reference collections are nearly essential (Fritz and Nesbitt 2014: 130–131). While charring helps to preserve some plant remains, it also modifies their shape. Archaeobotanists can conduct experiments to study the effects of heat on them, and it is important to include both charred and uncharred specimens in the reference collection (Hubbard and al-Azm 1990; Ucchesu et al. 2016).

16.6 Wood Charcoal of Trees and Shrubs

Dry-sieved sediment and heavy fraction from flotation often includes charcoal fragments, which field archaeologists also often pick out for radiocarbon dating or dendrochronology. These are not only useful for dating, but to reconstruct a site's environment and human preferences for fuel and construction materials (e.g., Byrne et al. 2013; Höhn and Neumann 2018). Even microscopic charcoal fragments from sediment samples can reveal the history of forest fires (Moore 2001).

However, taxa represented by charcoal may not be representative of the trees around a site because wood accumulates on the site through many cultural and natural processes that favor particular kinds of wood, while presenting many of the same taphonomic problems as faunal remains. Some kinds of wood make better fuel, are easier to hew, make straighter timbers, or are easily found lying on the ground within easy reach. Wood that is particularly useful can also be exported from forested regions to places with fewer or poorer trees, while charcoal can also come from driftwood that may have travelled thousands of kilometers on currents. In some instances, however, opportunistic collection of dead wood for fuel may provide a nearly random sample of wood in a site's vicinity. Only attention to site-formation processes and cultural practices can help us distinguish these situations (Théry-Parisot et al. 2010).

Accurate identification of wood or charcoal requires some preparation (Cutler et al. 2008; Stuijts 2006). It is necessary to break pieces so that they show the **transverse** (or cross-sectional), **radial**, and **tangential** planes (Fig. 16.10), using gentle pressure with a razor or scalpel to fracture, rather than slice or saw, the piece in the desired planes (Fig. 16.11). In the transverse section, in reflected light, the rays appear like spokes of a wheel, while in the tangential section, you see the rays end-on, and in the radial section they appear as parallel bands.

Identification depends on recognition of distinctive patterns in the cellular structure of the wood (Crang et al. 2018: 518–540; Crivellaro and Schweingruber 2013; Dubis n.d.; Helmling et al. 2018; Hoadley 1990; Lake 2015; Ruffinatto and Crivellaro 2019; Schweingruber 1990; Wheeler and Baas 1998). Identification to genus or species level requires detailed inspection of minute differences in cell structure, arrangement, and size. Consequently, a reference collection is essential (see Scheel-Ybert 2016). However, distinguishing hardwoods from softwoods more generally is less difficult (Fig. 16.10). Hardwoods have a complex structure with vessels (pores) running longitudinally through them; these appear as tunnel-like, circular features in transverse section. The vessels can be isolated or, as in maples and birches, grouped. Softwoods lack these specialized vessels, mainly showing the band-like variations in cell width (early and late growth) that occur in tree rings, although some genera, including pine, can have longitudinal resin canals that are somewhat similar to pores. They also show pits in radial section that are very useful in identification. Both kinds of wood show rays that radiate out in transverse section. In softwoods, the rays appear in tangential section as single or double strands of cells arranged vertically, but in hardwoods the rays can be multiseriate (have large bundles of cells arranged in vertical lenses). However, small fragment size and other factors sometimes makes it impossible to identify charcoal to one genus, requiring composite identifications, such as *Populus/Salix*.

In addition to taxonomic identification, careful examination of charcoal can inform us about the health of the original tree, branch diameter, and other factors relevant to firewood use and woodland management (Dufraisse 2006).

For reference collections, it is useful to mount thin sections cut from wood or charcoal on slides. This involves

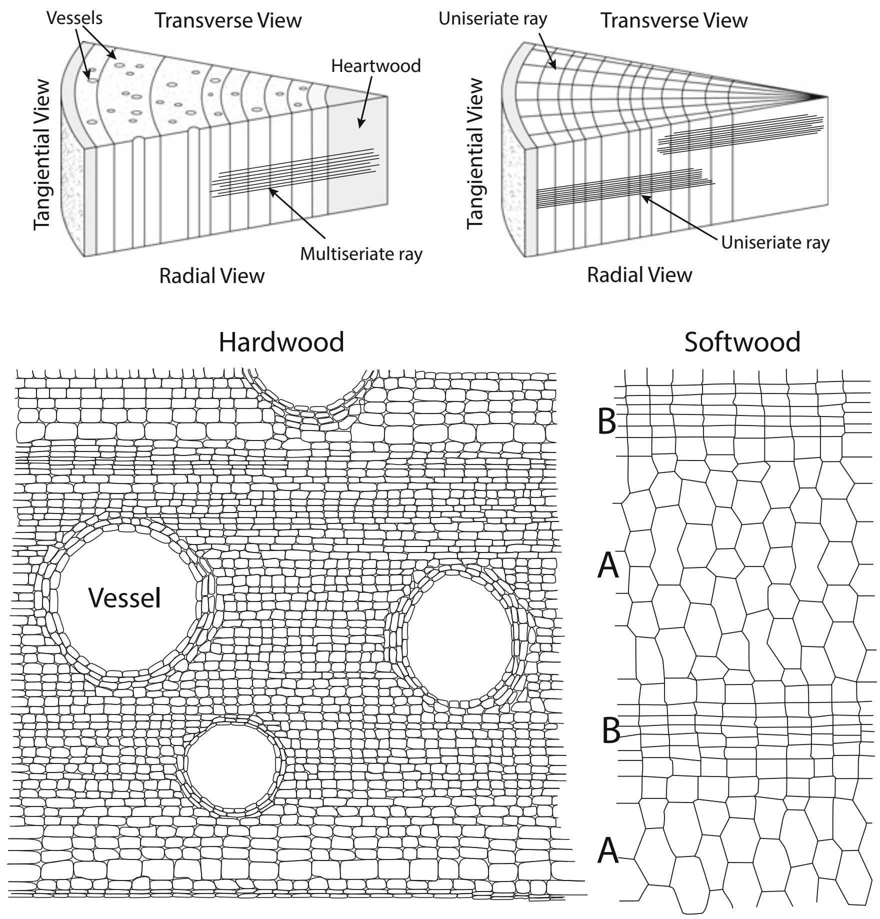


Fig. 16.10 Wedges of hardwood (top left) and softwood (top right) cut along transverse and radial planes, along with magnified transverse cross-sections to show the tubular vessels in hardwood (lower left), and the lack of vessels in softwood (lower right), which has early-wood (**a**) and late-wood (**b**) zones of growth. (Y. Salama)

Fig. 16.11 Fracturing a charred branch fragment in the transverse plane by gentle pressure with a sharp scalpel

cutting thin slices in each of the transverse, radial and tangential planes with a microtome or a very sharp razor blade, thin enough that light will show through. A dull blade will tear at the cell structure and ruin your section. You can mount the sections on a microscope slide in a solution of glycerin and alcohol. It is useful to keep one section of each plane on a single slide under three cover slips, with clear label and a consistent orientation and order from one to the other so that it is easy to switch views or slides without becoming disoriented. Heating the slide on a hot plate at 150 °C for 1–2 min expels air bubbles (Friedman 1978: 2–6).

Wood and its charcoal can also be useful chronologically because, in some species, seasonal and annual environmental variations affect the production of xylem and phloem to create tree rings of varying width (Crang et al. 2018: 516–518). This characteristic is the basis for dendrochronology (see chap. 20).

16.7 Parenchymous Remains

Vegetative remains of parenchymous plant organs (roots, tubers, rhizomes and corms) sometimes survive in archaeological deposits and can be particularly important in regions, like the tropics or Andes, where foods such as yams and potatoes have been staples (Pearsall 2019: 10–13).

Successful recovery of parenchymous tissues from sediments differs from that of seeds and nut fragments. Because they are fragile, machine-assisted water flotation is likely to break fragments into pieces too small for identification and, if water-flotation is necessary, it is better to use bucket flotation with only gentle agitation. Where it is practical, dry sieving is preferable, but without mechanical screen-shakers that would likely cause excessive fragmentation (Hather 2016: 74).

Because of the way they are processed, used, and discarded, large roots and tubers used for food are only likely to be charred and preserved as fragments, rather than whole organs. This presents challenges, and initial sorting by fragment types based on the kinds of tissues they preserve is a first step toward identification (Hather 2016: 72).

While charring or waterlogging can often aid their preservation, they may also distort the structure of parenchymous tissues by shrinking or swelling. Consequently, reference collections of parenchymous organs should be exposed experimentally to the same deteriorating conditions, such as charring, to improve the accuracy of identifications. Tissue that has dried slowly prior to charring shows the best preservation, while steam expulsion that results from heating moist tissues causes fissures and large vesicles (Hather 1991: 662).

Generally, the identification process for parenchymous tissues is similar to that for other plant remains, and especially charcoals, except that confident identification may require use of Scanning Electron Microscope (SEM), with reflected-light microscopy only serving as a first step (Fig. 16.12). As with charcoals, it is also necessary to cut or break a fresh, reasonably flat plane through each specimen (Hather 2016: 76).

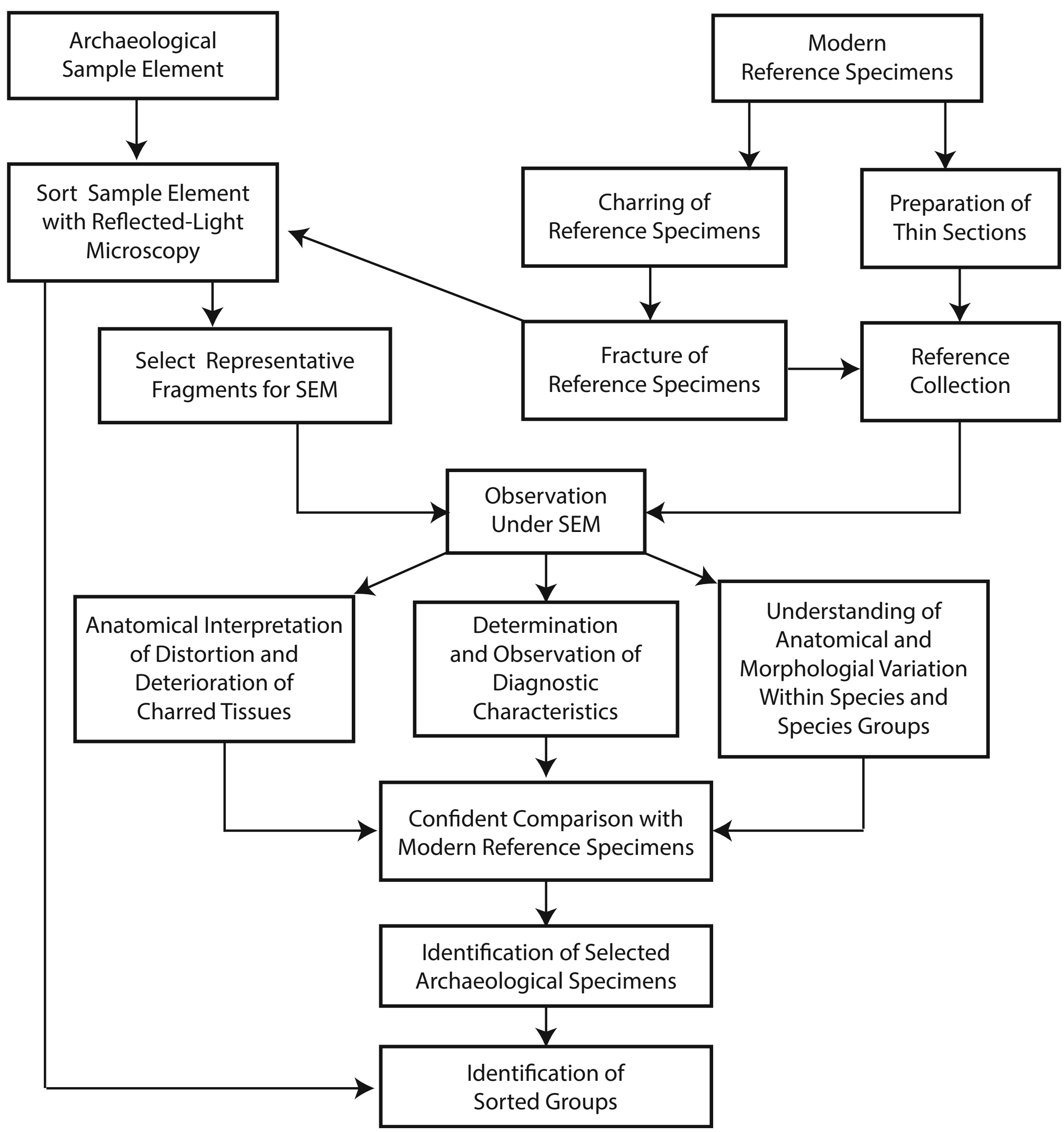


Fig. 16.12 Flow diagram for the identification of parenchymous tissues. (After Hather 2016: 75)

16.8 Processing and Analyzing Samples of Microremains

Microremains require use of separation and concentration techniques to isolate them from surrounding sediment, and high-power magnification for identification. They may also appear in sediment thin-sections.

The usual method for extracting phytoliths, starch grains and pollen from bulk sediments is called heavy-liquid flotation. This involves using a heavy liquid, such as lithium metatungstate, to separate particles of varying density (Chandler and Pearsall 2003; Coil et al. 2003; Horrocks 2005).

Micromorphology (see also pp. 302–303) involves removing blocks of sediment from select locations, consolidating and hardening them with resin, and sawing and polishing thin-sections of the blocks for examination under magnification. Sometimes phytoliths appear in these thin-sections and allow us to identify, for example, thin layers of degraded plant remains from straw stores, bedding, or the floors of livestock enclosures (e.g., Hubbard 2010).

16.8.1 Pollen

Archaeobotanists sometimes use pollen from archaeological contexts (Bryant and Hall 1993; Bryant and Holloway 1983) but off-site pollen records are particularly important as evidence for the vegetational environment and, indirectly, the climate of the region in which archaeological sites are located.

The most common context for recovering ancient pollen is in cores of waterlogged deposits, such as those in swamps and lake bottoms, because anaerobic environments favor pollen preservation and bodies of water are excellent for capturing wind-borne pollen. Pollen preservation is more favorable in sediments with pH no greater than 5.5, but there are other factors that also affect preservation (Havinga 1984).

It is usually necessary to concentrate pollen grains from sediment samples by progressive chemical digestion of the surrounding sediment (Bryant and Holloway 1983; Faegri and Iverson 1989: 72–84; Moore et al. 1991: 41–46). To aid quantification, analysts sometimes introduce a known number of polystyrene spheres, Lycopodium spores, or Eucalyptus pollen into fixed volumes of sediment (see p. 115). The addition of safranin red stain enhances pollen visibility. A portion of the concentrated pollen can then be mounted on a warm glass slide with a pipette, spread out, covered with a cover slip, and sealed with varnish. Once the slide is labelled, it is ready for examination under a microscope.

Identification of pollen grains depends on the great diversity of shapes that they exibit (Faegri and Iversen 1989;

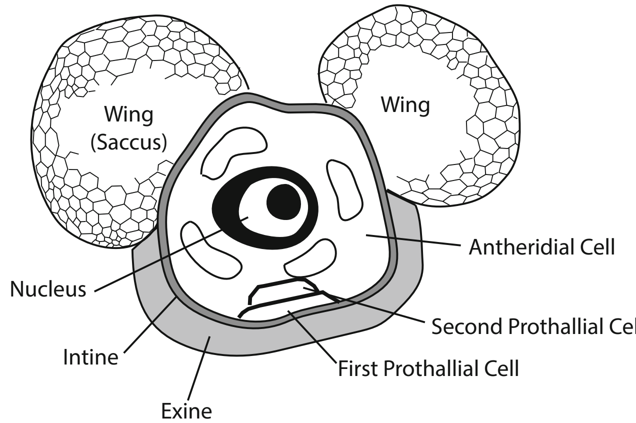


Fig. 16.13 View of a mature exine (decay-resistant outer layer) of Pine pollen, showing its parts

Moore et al. 1991). Both morphology and the relative abundance of pollen are related to the pollen's means of dispersal. **Anemophilous** (wind-borne) pollen, such as that of *Pinus*, *Casuarina*, *Poaceae*, and *Typha*, has features that enhance its buoyancy on the wind. For example, pine pollen has large sacs (Fig. 16.13) that make it very light (low in density). **Zoophilous** (animal-borne, including entomophilous or insect-borne) pollen comes from plants that produce less pollen but depend on animal vectors to transport it very effectively from the anthers to the stigmas of flowers. Pollen of this type has features that make it likely to attach to fine hairs on the animal meant to disperse it. **Hydrophilous** (water-borne) and cleistogamous (self-pollinating) pollens rarely occur in archaeological deposits because the former lacks a hard exine and the latter are not transported widely. Understanding pollen frequency distributions is impossible without considering the effects of these dispersal mechanisms (Hevly 1981).

Although there are guides for identification (e.g., Bassett et al. 1978; Moore et al. 1991; Weber 1998), as usual, there is no substitute for a voucher collection. These can be made from pollen collected in the field from confidently identified plants or from herbarium specimens; some are even available from commercial firms. However, pollen that is difficult to identify is typically only identified to a "type" (Greig 1989: 65).

Pollen analysis is a common source of evidence for changes in vegetation cover that might be related to climate change or human impacts. For example, the introduction of maize agriculture triggers relative decreases in tree pollen and increases in the pollen of field weeds (Fig. 16.14). Typically, palynologists examine changes in the pollen of several taxa of environmentally-sensitive trees and shrubs, as well as non-arboreal pollen—the collective contributions of all non-tree plants. Because of their interest in evidence for climate change and human impacts on landscapes, they can

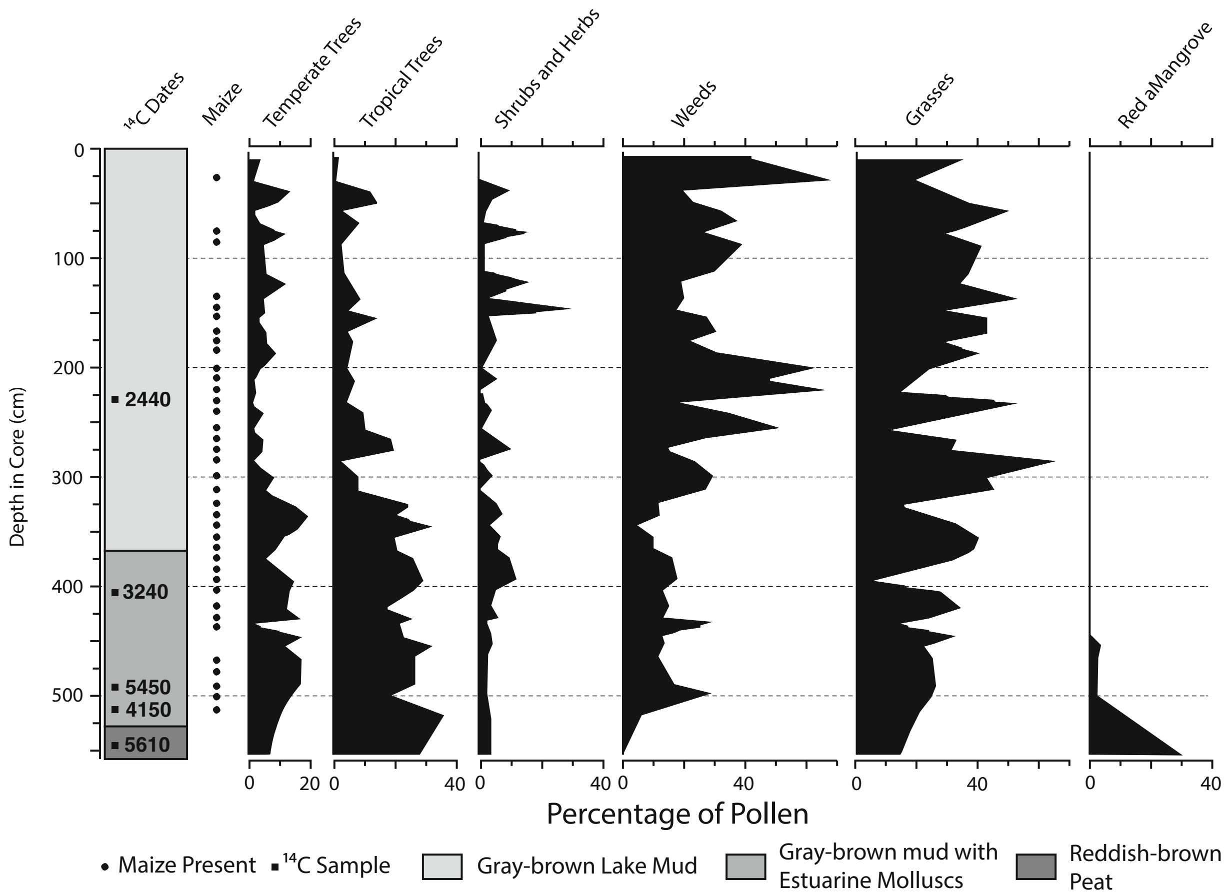


Fig. 16.14 Pollen diagram for a core from a lake in the coastal plain of Veracruz, Mexico, showing relative abundances for a variety of taxonomic groupings along with presence of maize, whose introduction is associated with a decrease in trees and increase in weed taxa. (Y. Salama, after Sluyter and Dominguez 2006). Radiocarbon determinations are uncalibrated and omit the errors

concentrate on groups of vegetation types that are sensitive to these effects because they are adapted to particular conditions of temperature and humidity. For example, tree cover may be more extensive during warmer, more humid climate, while certain shrubs and herbs that are adapted to steppes or prairies, such as *Artemisia sp.* (wormwood and sagebrush), may make stronger contributions to a pollen distribution when conditions are cooler and drier.

However, interpretation of pollen abundances requires careful consideration of plants' pollen productivity and the transport mechanisms that may have brought the pollen to the deposits that the cores intersected. Pollen that is easily carried on wind can travel many hundreds of kilometers, while zoophilous pollen might hardly be represented in a deposit even when there are thousands of these plants in its immediate vicinity. Consequently, the changing abundances of pollen taxa in a core do not necessarily reflect local changes in vegetation and may be better for interpreting regional-scale vegetation patterns. Interpreters of pollen evidence attempt to calibrate the varying kinds of pollen in light of these production and transport differences.

16.8.2 Phytoliths

The conditions under which phytoliths may be preserved in archaeological sediments are quite different from those for many other types of plant remains. Charring, for example, has negative impacts on phytolith preservation, while the presence of certain iron and aluminum oxides in sediments can enhance preservation. In addition, phytoliths of different taxa and even different plant parts may differ considerably in their stability over time (Cabanes and Shahack-Gross 2015; Cabanes et al. 2011; Pearsall 2014; Piperno 2006: 21–22).

Phytoliths pose unusual challenges of identification and quantification. The cell shapes that they document usually do not vary enough interspecifically to allow identification to species or genus, with rare possible exceptions, and also vary in different parts of the plant (Fig. 16.15). In addition, it is not obvious what counts of phytoliths would be telling us, as hundreds of phytoliths could come from a single stem fragment or from several fragments of unrelated plants. As with other kinds of plant remains, the sampled assemblage is a degraded remnant of a deposited assemblage that originated

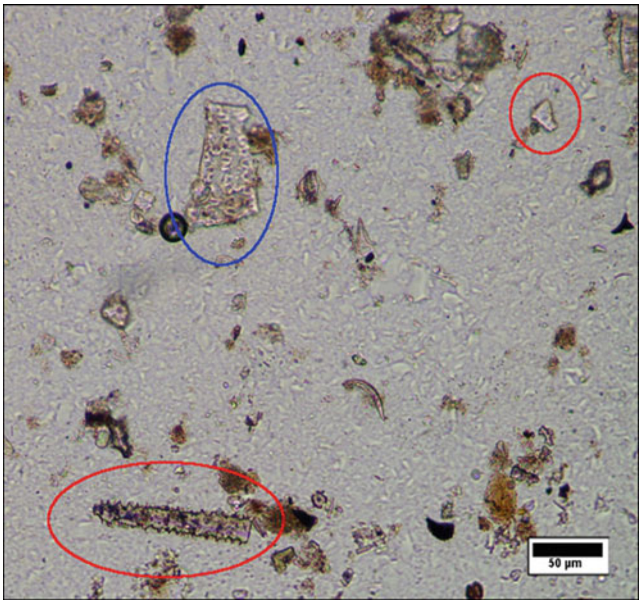


Fig. 16.15 Phytoliths (circled), with a saw-like grass phytolith at lower left. (Courtesy E. Hubbard)

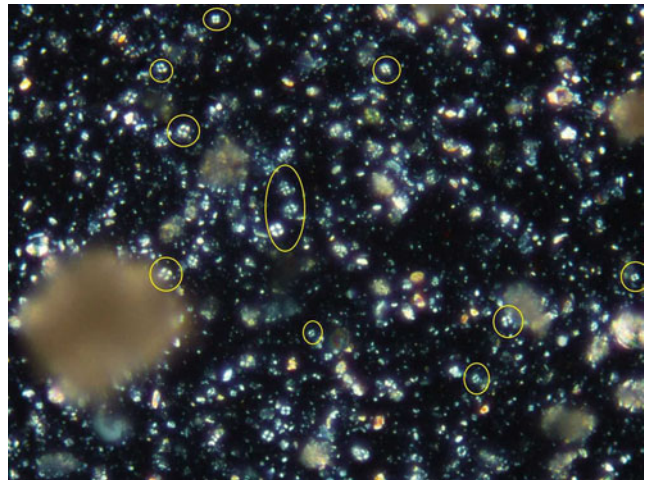


Fig. 16.16 Example of starch granules (circled) extracted from a Neolithic quern, each 5–10 µm in diameter, under polarizing microscopy. Note the cross-shapes that result from birefringence of the crystalline structures in starch. (Courtesy E. Yasui)

from behaviors that led to plant parts' accidental or intentional discard. Consequently, phytolith analysis tends to be most useful to archaeologists in the context of specific research questions for which phytoliths could provide supporting evidence.

Example
Alison Weisskopf (2017) exploits phytoliths of crop weeds in middens and other cultural deposits to distinguish between dependence on rice or millet agriculture in Neolithic and Early Bronze Age China. It was not possible or necessary to identify the taxa of the phytoliths, all of which came from grass leaves, but the distinction between phytoliths that came from cells genetically predisposed to form phytoliths and those from cells that form phytoliths only under wet conditions allowed her to distinguish wet and dry cycles. Millet could be grown under dry conditions, but rice would only be grown when there was sufficient water.

16.8.3 Starches

Although they are far from immune from degradation or decay (Haslam 2004; Henry 2014), starch granules can sometimes survive for many millennia in sediments, on tool and pottery surfaces, and in dental calculus. The crystalline layer in starch granules causes a distinctive, **birefringent** "Maltese Cross" when viewed under the microscope (Fig. 16.16; Copeland and Hardy 2018; Torrence and Barton 2006), and starch granules also exhibit internal growth rings. Because they vary substantially in size, shape, and molecular structure (Reichert 1913; Torrence et al. 2004), both among species and among plant organs, they provide a valuable complement to other plant evidence, and particularly for plant foods, such as tubers, that are not consistently represented among macroremains. When starch is cooked in water, however, the granules swell and lose their distinctive molecular organization, while processing of foods by grinding and pounding can also damage granules (Lamb and Loy 2005).

Starch analysis requires removal from artifacts, coprolites or dental calculus and separation from soil and other minerals. Sometimes, previous cleaning of artifacts has removed most residues, but starch can sometimes still be present in pores and crevices. Removal methods vary, and include extraction by pipette and distilled water, mechanical removal with sterile tools, washing with water and ultrasonic cleaning. The enzyme, alpha-amylase, can confirm that starch is present, while repeated agitation and centrifuging in a solution of cesium choloride (CsCl, $\rho = 1.8$ g.cm^{-3}) and washing with distilled water can separate the starch (e.g., Pagán-Jiménez et al. 2015). A droplet of the more concentrated starch solution is placed on a new slide with a small amount of glycerol to enhance the birefringence of the crystalline layers. Staining, often with CongoRed or Trypan Blue, can also be useful, particularly for identifying damaged grains. For longer-term storage, a cover glass will protect the specimen on the slide to be rehydrated for later analysis after it dries out (Piperno 2006), or it can be mounted for a permanent collection with glycerol, immersion oil, or similar compounds.

As with other classes of plant remains, identification depends a good deal on the availability of a modern reference collection, although such collections are currently fairly rare. Starch analysis is still in its early stages, and there remains skepticism about the reliability of starch identifications that are based solely on visual examination under the microscope (Akeju et al. 2018).

Other key challenges in starch research are taphonomy and controls for contamination (Barton and Matthews 2006; Barton and Torrence 2015). As with other kinds of plant evidence, it is useful to think in terms of the processes that are more or less likely to result in deposition of starch.

16.9 Chemical and Isotopic Evidence

Chemical residues on the edges of cutting tools or on the use surfaces of grinding stones can provide some of our best evidence for their use. Not only can these sometimes show phytoliths, pollen, or starches, as noted above, but also cellulose and lignin (polymers that give rigidity to cell walls, bark and wood), lipids (molecules in fats and oils), resins, or amino acids. The methods used to detect and identify these residues include gas chromatography (GC), mass spectrometry (MS), a very useful combination of these (GC-MS), gas chromatography-combustion-isotope ratio mass spectrophotometry (GC-C-IRMS), Fourier Transform Infrared Spectroscopy (FTIR), and others. Much of this research has focussed on identification of animal fats (see pp. 257–258), but they have also been successful at identifying proteins and lipids from plants (e.g., sesame oil, Shevchenko et al. 2017).

Isotopic research on the plant component of diet has focussed mainly on the stable carbon istopes, ^{13}C and ^{12}C (as opposed to radioactive ^{14}C). The ratio of these isotopes in living organisms varies because the photosynthetic pathways of plants differ in how they take them up from carbon dioxide in the atmosphere. This slight preference for one isotope over another is called fractionation, and archaeological analysis of these isotopes depends on differences between two groups of plants with different photosynthetic pathways, C3 and C4 plants. The former group includes trees, shrubs, and temperate grasses; the latter includes tropical grasses, such as maize, and tends to have more ^{13}C relative to ^{12}C. Consequently, when humans consume substantial amounts of maize or other C4 plants, you would expect that to show up in the isotopic composition of the carbon in their bones (Fig. 16.17). Because humans are omnivores, the mixing of isotopic signals in their diets can be very complex, and some researchers are beginning to use Bayesian methods (see pp. 139–140) to try to sort that out (Lewis and Sealy 2018).

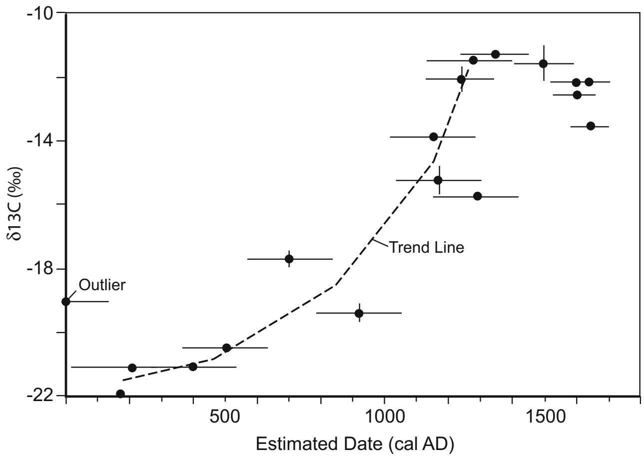


Fig. 16.17 Time-series graph showing variation in the mean ^{13}C to ^{12}C ratio ($\delta^{13}C$ in ‰) at sites in southern Ontario. (After Katzenberg et al. 1995, error bars estimated by author). Note the rather rapid change in this ratio about 1000 cal AD, marking the introduction of maize as a staple in diet. As the points are means by site, they do not reflect all the variation in the data

16.10 Quantification Issues

As already discussed at length in Chap. 7, quantification is a thorny issue for archaeobotanists. We can count seeds or pollen grains, or measure the mass of charcoal fragments, but what do the resulting measures mean in terms of plant use, environment, or past cultures?

Seeds, glumes, internodes, rachis fragments, and many other plant fragments are amenable to counting, and it is sometimes possible to avoid double-counting of fragments by only counting those that are more than 50% complete or those that preserve a "landmark," such as the hilum on a dichot seed. Differential fragmentation makes the interpretation of charcoal or parenchymous counts very difficult (Chrzazvez et al. 2014), and most researchers prefer to quantify these by mass.

However, there has long been controversy over how we can make use of counts or mass for valid interpretation. The appropriate choice depends on the research questions and the taphonomic processes to which assemblages were probably exposed.

One approach has been to make use of ratios, especially the ratio of charred seed counts to mass of charcoal (pp. 112–113; Marston 2014; Miller 1988: 75–83). The idea here is that charcoal mass normalizes for at least some of the preservation issues, but this depends on the questionable assumptions that habitual use of wood as fuel is rather constant and that preservation rates of charcoal and other charred plant remains are very similar and stable (Kadane 1988: 212; see Orton 2000: 65–66).

Rather than take their absolute magnitudes or their proportions very seriously, many analysts are interested in the implications of various combinations of macroemains for the character of the deposit in which they occur and the processing or discard activity with which they are most likely associated. For example, Hillman's work in Turkey, as described earlier in this chapter, suggests that a plant assemblage in which spikelet forks and tail grains from glumed wheats are fairly common, while glume bases and internodes, and weed seeds smaller in size than prime grain are very abundant, probably resulted from fine sievings during grain processing. The exact quantities of these things are not important, only their ordinal-scale representation and the way they are combined with other plant remains as evidence for the processing stage that could have resulted in a deposited assemblage. These characterizations are also most useful in the context of specific research questions.

Similarly, we can study the distributions of wood charcoal in the context of how people selected wood for fuel or lumber. Fuel wood, of course, is much more likely to be preserved by charring, except in the case of catastrophic or intentional fires that engulfed whole buildings. Where we use charcoal as evidence for the environments around sites, it may be useful to compare the archaeological charcoal distributions, usually quantified by mass, with the distributions of tree taxa in modern environments (Smart and Hoffman 1988 184). However, selection of wood for household fuels is rarely completely random, since different woods vary in their ease of collection and burning qualities, and fuel suitable for some industrial processes, such as smelting ore or firing pottery, could be even more selective or include imported fuel with no relationship to local forests (Smart and Hoffman 1988:168–169). Another quantitative aspect of archaeoanthracology (charcoal analysis) is to estimate the mean diameter of the wood from which charcoal fragments originated. This is biased unless bark is routinely present and all fragments are large enough to measure the curvature of growth rings and angles between rays, but it still gives a sense of whether fuel consisted of small branches and coppice shoots, larger branches, or thick portions of tree trunks (e.g., Jansen and Nelle 2014). It can also be useful for estimating the relative merits of fragments for dendrochronology or radiocarbon dating (see Chap. 20).

Pollen and similar microremains have their own quantification challenges. Counting pollen involves making a series of traverses at about 400x across the slide by turning the small "travel knobs" to move the microscope stage in the x- and y-axes by small, fixed increments. To avoid confusion, always begin at the same corner of the slide. Along each traverse, you count and record pollen of various taxa at various grid coordinates. Grains that are not immediately identifiable require closer examination by zooming in at higher magnification, perhaps 1000x (Faegri and Iversen 1989: 65–66).

Pollen quantification is often expressed as number of grains per unit volume or unit mass of sediment (as measured before chemical digestion) rather than simply in number of grains or percentage of grains, although the last is still common. Pollen analysts can use a sampling-resampling approach (p. 115) by placing a known number of "exotics" (microremains that would not occur naturally), such as *lycopodium* (clubmoss) spores in the sample volume. The proportion of "exotics" provides a basis for calibrating the number of pollen grains and providing a reliable estimate for the total number of ancient pollen of each taxon in that sample element (Benninghoff 1966). For example, if you know that you placed 1000 *lycopodium* spores in a volume of sediment that should not have had any *lycopodium* in it, and then find 50 *lycopodium* spores in the sub-sample that you examined, you can multiply the counts of all the ancient pollen grains by 20 (1000/50) to obtain an estimate of the total number of each type of pollen in the volume.

For starches and chemical residues, it is not necessary to count anything. Instead, the mere presence of these residues—after controlling for possible contamination—provides evidence for contact with or consumption of certain categories of plants.

16.11 Archaeobotany and Paleoenvironment

Archaeobotanical evidence, especially from off-site pollen and archaeological wood charcoal, is important for reconstructing prehistoric environments and the place of humans in them (Barker 1985: 15–19; Crawford 2018: 156–158).

A key archaeological interest is interaction of humans with vegetation communities in their habitats. A fairly early paradigm for such research is human ecology (Bruhn 1974; Butzer 1982; Hardesty 1977; Lawrence 2003; Vayda 1969), along with the closely related historical ecology (Crumley 1994; Szabó 2015). Although human ecology is far from being a unified paradigm, it features humans as just one species among many in an ecosystem that involves interspecific exchanges of matter, energy and information. It focuses specifically on the interrelationships between human populations and the other components, both living and non-living, in ecosystems. While archaeobotany can play a large role in human ecology, it is inherently interdisciplinary.

Human Behavioral Ecology (HBE), more a branch of human ecology than a completely different paradigm, is less common in archaeobotany. Like human ecology more generally, it contextualizes human behaviors, including those involving plants and animals, within evolutionary ecology. However, with its connections to ecology and economics, HBE makes use of decision theory and other optimizing functions to determine the "best" combinations of resource use in terms of costs and benefits, as measured in energy or some other "currency" (e.g., Gremillion 2014; Winterhalder and Smith 2000).

Niche-Construction Theory (NCT) has recently become common in archaeobotany. Odling-Smee (1988; Odling-Smee et al. 2003) introduced the term niche construction to describe an evolutionary process that Lewontin (1983) had already been advocating as a complement to natural selection (Laland et al. 2016). A key feature of niche construction theory is that an organism does not simply adapt to pre-existing conditions through selection, but can itself modify its environment in ways that influence the selection pressures on itself or other organisms so that there is an evolutionary response in at least one species (Matthews et al. 2014). While niche construction itself is fairly obvious—all organisms have some kind of impact on their environments—the difference in NCT is that these impacts alter the nature or intensity of natural selection for one or more organisms. It also tends to be a reciprocal relationship, as adapting organisms can make further alterations to their environments as they respond to an earlier one.

Archaeologists are of course most interested in the implications of NCT for human interactions with their environments, and this is particularly true of research on domestication (see below). Bruce Smith (2012, 2014) argues that NCT provides a much better explanation for plant (and animal) domestication than HBE or human ecology more generally can do, particularly as it does not require the assumption that humans were adapting to adverse climatic changes or population growth that restricted resource availability. We can expect that many of the resource-collecting behaviors and mobility decisions of Pleistocene and early Holocene hunter-gatherers to have had substantial impacts on the distribution and density of both desirable and undesirable plants. Some of these impacts, furthermore, would have been ones that altered selective pressures on the plants, on humans, or on other animals that preyed on the plants. Smith (2012: 264) is particularly interested in humans' deliberate modifications to their environments to enhance availability of desirable resources, but it is important to note that both deliberate and unintentional modifications can satisfy the criteria for niche construction. For example, humans may deliberately till a field to increase their yield of some crop, while unintentionally creating conditions that change the selective pressures on weeds or non-human animals that also prey on that crop.

16.12 Reconstructing Paleodiets, Foodways and Plant Use

As in zooarchaeology, archaeobotanists place considerable emphasis on the role of plants in ancient dietary choices and increasingly in culinary practices and social relations (Hastorf 2017; Morehart and Morell-Hart 2015). They are interested not only in staple foods, but in plant use for fuel, construction or clothing material, drugs, condiments, spices, cosmetics, dyes, inks, and poisons (Day 2013; Pearsall 2019: 198–227).

As noted in Chap. 7 and in a previous section, interpreting the quantitative evidence for plant use is complicated, and abundances of plant remains do not just translate into dietary preferences, niche breadths or culinary practices. It is easier to identify the introduction of new plants or their adoption as staples, while the newer research on starch and chemical residues is increasingly useful for identifying actual food-preparation practices. Evidence from dental calculus, in particular, provides our most direct evidence of what people were eating, although with less specificity than we would typically like.

Example

To study diet and its cultural implications, Beck et al. (2016) examine changes in the distributions of plant and animal foods from earlier to later contexts at the sixteenth-century, Spanish colonial site of Fort San Juan de Joara in what is now North Carolina. Counts of charred acorns, hickory nuts, maize, and other food crops and fruit from the Spanish compound just north of the fort strongly suggest that the Spanish soldiers garrisoned there depended on indigenous women for their plant food provisions, rather than provisions brought from elsewhere or crops from imported seed stocks. Statistical analyses show significant differences between early and later contexts in the compound, mainly involving a shift from acorn to hickory that may be due to local women adapting to the Spanish soldiers' preferences or to the addition of nonlocal slaves to the labor pool. Evidence from animal remains similarly shows a shift from early contexts in which deer and bear contributed almost equally to the soldiers' diet to later ones in which venison—arguably more sympathetic to European palates—became by far the dominant meat in the Spanish compound and became less likely to come from whole deer carcasses. As in the case of the plant foods, this could be due to local hunters adapting to Spanish demand.

16.13 Plant Domestication and the Origins of Food Production

Archaeobotanists have long focused attention on identifying and understanding the domestication of plants and the dispersal of agricultural ecologies. Domestication is a process whereby a plant, through natural or artificial selection, or both, becomes dependent on humans for its dispersal just as humans become more dependent on the plant, or one that grows with it, for food or some other use. Crop plants and the weeds that accompany them are both subject to these evolutionary influences (Willcox 2012).

The domestication of wheat, barley, rice, maize, squash, millet and pulses have attracted a great deal of interest (Bestel et al. 2014; Conard 2016; Crawford 2016; Fuller 2007; Fuller et al. 2011; Gross 2012; Milla et al. 2015; Ranere et al. 2009; Sang and Ge 2007; Sonnante et al. 2009; Wang et al. 2005; Willcox 2013; Zheng et al. 2016). However, there has also been interest in the domestication of figs, olives and grapes in the Old World (e.g., Denham 2007; Gismondi et al. 2016; Kislev et al. 2006; Margaritis 2013; McGovern et al. 2017) and of potatoes in the Andes (Spooner et al. 2005). Not surprisingly, studies of domestication have tended to have a human-ecological or evolutionary focus (e.g., Milla et al. 2015; Punugganan and Fuller 2009; Rindos 1984), and increasingly a niche-construction one (e.g., Smith 2012), while genetics has had a profound impact on our understanding over the last two decades (e.g., Kovach et al. 2007; Peleg et al. 2011; Pourkheirandish et al. 2015).

Domestication is related to, but distinct from, the human behaviors that involve manipulating and tending plants; in some instances, humans may have been cultivating wild plants for decades or centuries before the plants showed any morphological signs of domestication (e.g., Willcox et al. 2008), although domestication processes can also be rapid. Other kinds of evidence can sometimes allow us to infer behavioral changes associated with cultivation, whether or not morphological changes in crop plants have yet developed, such as the appearance or shift in abundance of weed taxa that typically appear only in assemblages harvested from cultivated fields (Hillman 2000: 384–388).

Some, but not all, of the following changes that plants often undergo in the domestication process are morphological, making them useful evidence for domestication (Rindos 1984: 183):

- The plant loses its ability to disperse seeds. For example, wild grasses have a brittle rachis that shatters when the grain is ripe so that grain will scatter; domestic wheat has a tough rachis that holds grain until after humans harvest it
- The plant part that humans consume becomes larger. In some cases, such as apples, it is not only enlarged, but attracts humans and other predators by signaling ripeness with color changes
- The plant part that humans use may become clustered. Seeds that are clustered in a pod, spike or maize cob are easier to harvest in large quantities, a feature that attracts predators, including humans
- There is often a change in duration, from annual to perennial, or the reverse
- There is a tendency toward polyploidy (increase in chromosome number), often accompanying gigantism
- There is a loss of dormancy. In the wild, only some seeds germinate in their first year, while others lie dormant until later years as a sort of insurance against drought or other poor growing conditions. In an agricultural ecology, where humans seed and tend the plants, this feature usually disappears
- Plants tend to develop simultaneous ripening. In the wild, fruit may ripen at different times over several weeks, again as a sort of insurance against bad weather or predators. Simultaneous ripening is attractive to humans because they can harvest a crop all at once
- Plants tend to lose features, such as thorns and toxins, that protect wild plants from predation. Humans would tend to select less thorny and more palatable plants for their use, and may aid their dispersal and increase, while plants in fields tended by humans have less need for these protective features
- Diversity in plant form tends to increase. Subtle mutations that would not have persisted in the wild may thrive in the protective environment of humans' fields, and humans may even encourage this diversity by intentional propagation
- Because humans provide a dependable means of reproduction, plants tend to change from very opportunistic (r-selected) to more specialized (K-selected) species during the domestication process.

While early domestication events have dominated the literature, domestication is an ongoing evolutionary process. New domesticates can be added to agricultural ecologies long after introduction of the earliest domesticates, while existing domesticates can develop new varieties or even species through selection and hybridization.

Case Study
Plant Remains from Sites in Korea

Gyoung-Ah Lee (2012) provides an example of how archaeobotanists can grapple with the dual problems of taphonomic effects and effective sample size. In large, semi-subterranean houses at several sites of the Early Chulman through Late Mumun periods (ca. 5600–200 BC) in the Nam River valley, Korea, Lee was interested in identifying changes from a foraging to an agricultural economy. The most significant changes in this sequence probably occurred from the Early to Middle Mumun. Given this research question, she hoped to characterize plant remains from whole sites that dated to particular intervals over this long period, rather than, for example, distinguishing differences in plant use among houses or features.

However, the potential numbers of plant remains to analyze were huge and unknown taphonomic effects that had probably affected the sampled assemblages differentially were major challenges.

In an attempt to deal with these problems, while still allowing her to answer her research questions with reasonable confidence, Lee attempted to determine what sample sizes she would need to analyze from the hundreds of sample elements that had been collected from these sites. The sample elements were 10 L of sediment from each 1 m^2 grid square and from a subsample of pit features. Taking the total of these at each site as the population, and given the substantial labor costs of counting seeds in such volumes after flotation, Lee estimates the sample size (number of grid squares and features) necessary to achieve a relative standard error (RSE) on seed densities of 20% at 90% confidence ($t = 1.83$ and $r = 0.2$) for each period of occupation. She assumes that these densities will provide a sound basis for estimating the proportions of taxa among the seeds of whole sites. The result was $n = 141$ for the Early Mumun contexts, and $n = 70$ for the Middle Mumun. However, she was actually able to accomplish her goals with even smaller sample sizes, as samples of $n = 76$ and n = 63 turned out to meet her objective of 20% RSE at 90% confidence. This strategy provided a basis for inferences at the site level without the cost of having to analyze 10 L volumes from all of the grid squares and features.

She then compares the samples by using the ratio-of-ratios approach (Orton 2000: 65–66). This makes comparisons among sites, contexts or time periods possible when we can plausibly assume that taphonomic differences are such that they still preserve proportionality among taxa (see p. 119 and Fig. 7.3). For example, if Millet was twice as abundant as rice at one site and four times as abundant at another (ratio of 2:1 between sites), that same ratio will be found in the degraded sample that the archaeobotanist analyzes (still 2:1), even though the two sites' preservation environments differ, as long as the probability of survival of millet and rice retains proportionality (e.g., 0.8 to 0.4 at one site and 0.4 to 0.2 at the other).

16.14 Quality in Archaeobotanical Analysis

The previous chapter ended with an example of interobserver studies to evaluate the quality of zooarchaeological identifications. Analyses of plant remains are similarly vulnerable to errors in identification, quantification and interpretation, as well as recovery (e.g., Hosch and Zibulski 2003).

For example, Wright (2005) argues that variations in the ways that different archaeobotanists collect and process sediments for flotation probably have a large impact on quantification and resulting interpretations. She uses an experiment with three 10 L volumes of sediment (sandy loam, silt-loam, and clay-loam) that each contained exactly 25 charred specimens of each of 11 taxa of charred nutshell, wood, and seeds to estimate their recovery rates in both the heavy and light fractions that result from use of a Shell Mound Archaeological Project (SMAP) flotation system, which agitates with water jets, and 1 mm mesh insert. Wright's findings indicate that, while recovery of charcoal and nutshell was fairly good, recovery of other remains was highly variable and especially poor for small seeds, such as tobacco and chenopod. Clearly, this recovery variability has major implications for the value of reported abundances of plant remains and potentially even for measures of ubiquity and diversity when some taxa are relatively rare. Furthermore, this study does not account for variability among flotation machines or analysts.

16.15 Summary

- Plant evidence consists of macroremains (e.g., charred endocarps and charcoal), microremains (e.g., pollen, starches, and phytoliths), and chemical and isotopic traces
- Differential preservation and recovery are major concerns for archaeobotanists. Orton's ratio-of-ratios approach is a creative attempt to account for this

- Archaeobotany has long made important contributions to the study of human-environmental interactions, sometimes in a human ecological framework, and recently with increasing emphasis on niche construction theory (NCT)
- Archaeobotany has also made contributions to understanding dietary choices and plant use in the past, although the quantitative problems have been challenging. Today, isotopic methods, starch residues and other sources of evidence have provided new and effective tools for this work
- Other major foci of archaeobotany have been agricultural origins, food processing, plant domestication, and the evolution of agricultural ecologies

References Cited

Akeju, T., Brown, M., Bundala, M., Collins, M. J., Copeland, L., Crowther, A., et al. (2018). Exaggerated expectations in ancient starch research and the need for new taphonomic and authenticity criteria. *Facets*. https://doi.org/10.1139/facets-2017-0126.

Barker, G. (1985). *Prehistoric farming in Europe*. Cambridge: Cambridge University Press.

Barton, H., & Matthews, P. J. (2006). Taphonomy. In R. Torrence & H. Barton (Eds.), *Ancient starch research* (pp. 75–94). Walnut Creek: Left Coast Press.

Barton, H., & Torrence, R. (2015). Cooking up recipes for ancient starch: Assessing current methodologies and looking to the future. *Journal of Archaeological Science, 56*, 194–201.

Bassett, I. J., Crompton, C. W., & Parmalee, J. A. (1978). *An atlas of airborne pollen grains and common fungus spores of Canada*. Ottawa, ON: Department of Agriculture.

Beck, W. (1989). The taphonomy of plants. In W. Beck, A. Clarke, & L. Head (Eds.), *Plants in Australian archaeology* (pp. 31–53). (Tempus, archaeology and material culture studies in archaeology 1). St. Lucia, QLD: University of Queensland.

Beck, R. A., Fritz, G. J., Lapham, H. A., Moore, D. G., & Rodning, C. B. (2016). The politics of provisioning: Food and gender at Fort San Juan de Joara, 1566-1568. *American Antiquity, 81*(1), 3–26.

Benninghoff, W. S. (1966). Calculation of pollen and spore density in sediments by addition of exotic pollen in known quantities. *Pollen et Spores, 44*, 332–333.

Bestel, S., Crawford, G. W., Liu, L., Shi, J., Song, Y., & Chen, X. (2014). The evolution of millet domestication, middle Yellow River region, North China: Evidence from charred seeds at the late upper Paleolithic Shizitan locality 9 site. *The Holocene, 24*(3), 261–265.

Bruhn, J. (1974). Human ecology: A unifying science? *Human Ecology, 2*(2), 105–125.

Bryant, V. M., & Hall, S. A. (1993). Archaeological palynology in the United States: A critique. *American Antiquity, 58*(2), 277–286.

Bryant, V. M., & Holloway, G. (1983). The role of palynology in archaeology. *Journal of Archaeological Method and Theory, 6*, 191–224.

Butzer, K. W. (1982). *Archaeology as human ecology*. Cambridge: Cambridge University Press.

Bye, R. A. (1986). Voucher specimens in ethnobotanical studies and publications. *Journal of Ethnobiology, 6*, 1–8.

Byrne, C., Dotte-Sarout, E., & Winton, V. (2013). Charcoals as indicators of ancient tree and fuel strategies: An application of anthracology in the Australian Midwest. *Australian Archaeology, 77*, 94–106.

Cabanes, D., & Shahack-Gross, R. (2015). Understanding fossil phytolith preservation. The role of partial dissolution in paleoecology and archaeology. *PLoS ONE, 10*(5), e0125532.

Cabanes, D., Weiner, S., & Shahack-Gross, R. (2011). Stability of phytoliths in the archaeological record: A dissolution study of modern and fossil phytoliths. *Journal of Archaeological Science, 38*, 2480–2490.

Chandler-Ezell, K., & Pearsall, D. (2003). 'Piggyback' microfossil processing: Joint starch and phytolith sampling from stone tools. *The Phytolitharien, 15*, 2–8.

Chiou, K. L., Cook, A., & Hastorf, C. (2013). Flotation versus dry sieving archaeobotanical remains: A case history from the Middle Horizon southern coast of Peru. *Journal of Field Archaeology, 28*, 38–53.

Chrzazvez, J., Théry-Parisot, I., Firucci, G., Terral, J., & Thibaut, B. (2014). Impact of post-depositional processes on charcoal fragmentation and archaeobotanical implications: Experimental approach combining charcoal analysis and biomechanics. *Journal of Archaeological Science, 4*, 30–42.

Coil, J., Korstanje, A., Archer, S., & Hastorf, C. (2003). Laboratory goals and considerations for multiple microfossil extraction in archaeology. *Journal of Archaeological Science, 30*, 991–1008.

Conard, N. (2016). The origins of agriculture in South-West Asia. In N. Sanz (Ed.), *Origins of food production* (pp. 50–67). Mexico City: UNESCO Office Mexico.

Copeland, L., & Hardy, K. (2018). Archaeological starch. *Agronomy, 8* (4). https://doi.org/10.3390/agronomy8010004.

Crang, R., Lyons-Sobaski, S., & Wise, R. (2018). *Plant anatomy*. Cham, CH: Springer.

Crawford, G. W. (2016). Trajectories to agriculture: The case of China, Japan and the Korean peninsula. In N. Sanz (Ed.), *Origins of food production* (pp. 102–113). Mexico City: UNESCO Office Mexico.

Crawford, G. W. (2018). Palaeoethnobotanical contributions to human-environment interaction. In E. Pişkin, A. Marciniak, & M. Bartkowiak (Eds.), *Environmental archaeology: Current theoretical and methodological approaches* (pp. 155–180). New York: Springer.

Crivellaro, A., & Schweingruber, F. H. (2013). *Atlas of wood, bark and pith anatomy of eastern Mediterranean trees and shrubs, with a special focus on Cyprus*. New York: Springer.

Crumley, C. L. (1994). Historical ecology: A multidimensional ecological dimension. In C. L. Crumley (Ed.), *Historical ecology: Cultural knowledge and changing landscapes* (pp. 1–16). Santa Fe: School of American Research.

Cutler, D. F., Botha, C. E. J., & Stevenson, D. W. (2008). *Plant anatomy: An applied approach*. Malden: Blackwell.

Day, J. (2013). Botany meets archaeology: People and plants in the past. *Journal of Experimental Botany, 64*(18), 5805–5816.

Delorit, R. J. (1970). *Illustrated taxonomy manual of weed seeds*. River Falls, WI: Agronomy Publications.

Denham, T. (2007). Early fig domestication, or gathering of wild parthenocarpic figs? *Antiquity, 81*(312), 457–461.

Dimbleby, G. W. (1967). *Plants and archaeology*. London: John Baker.

Dubis, J. (n.d.). *Identification of common North American woods*. https://woodidentification.net/contact/. Accessed 31 Dec 2018.

Dufraisse, A. (2006). Charcoal anatomy potential, wood diameter and radial growth. In A. Dufraisse (Ed.), *Charcoal analysis: New analytical tools and methods for archaeology* (BAR international series 1483) (pp. 47–59). Oxford: Archaeopress.

Evershed, R. P., Heron, C., Charters, S., & Goad, L. J. (1992). The survival of food residues: New methods of analysis, interpretation, application. *Proceedings of the British Academy, 77*, 187–208.

Faegri, K., & Iverson, J. (1989). *Textbook of pollen analysis* (4th ed.). Oxford: Blackwell.

Farahini, A., Chiou, K. L., Harkey, A., Hastorf, C. A., Lentz, D. L., & Sheets, P. (2017). Identifying 'plantscapes' at the classic Maya village of Joya de Cerén, El Salvador. *Antiquity, 91*(358), 980–997.

Friedman, J. (1978). *Wood identification by microscopic examination: A Guide for the archaeologist on the Northwest Coast of North America* (British Columbia Provincial Museum heritage record 5). Victoria, BC: Province of British Columbia Ministry of the Provincial Secretary and Travel Industry.

Fritz, G., & Nesbitt, M. (2014). Laboratory analysis and identification of plant macroremains. In J. M. Marston, J. d'A. Guedes, & C. Warinner (Eds.), *Method and theory in paleoethnobotany* (pp. 115–145). Boulder, CO: University Press of Colorado.

Fuller, D. Q. (2007). Contrasting patterns in crop domestication and domestication rates: Recent archaeobotanical insights from the Old World. *Annals of Botany, 100*(5), 903–924.

Fuller, D. Q., Willcox, G., & Allaby, R. G. (2011). Cultivation and domestication had multiple origins: Arguments against the core area hypothesis for the origins of agriculture in the near east. *World Archaeology, 43*(4), 628–652.

Gallagher, D. E. (2014). Formation processes of the macrobotanical record. In J. M. Marston, J. d'A. Guedes, & C. Warinner (Eds.), *Method and theory in paleoethnobotany* (pp. 19–34). Boulder, CO: University Press of Colorado.

Gismondi, A., Di Marco, G., Martini, F., Sarti, L., Crespan, M., Martínez-Labarga, C., et al. (2016). Grapevine carpological remains revealed the existence of a Neolithic domesticated *Vitis vinifera* L. specimen containing ancient DNA partially preserved in modern ecotypes. *Journal of Archaeological Science, 69*, 75–84.

Greig, J. (1989). *Archaeobotany (handbooks for archaeologists 4)*. Strasbourg: European Science Foundation.

Gremillion, K. J. (2014). Human behavioral ecology and paleoethnobotany. In J. M. Marston (Ed.), *Method and theory in paleoethnobotany* (pp. 339–354). Boulder: University Press of Colorado.

Gross, B. L. (2012). Rice domestication: Histories and mysteries. *Molecular Ecology, 21*(18), 4412–4413.

Hardesty, D. L. (1977). *Ecological anthropology*. New York: Wiley.

Harlan, J. R. (1967). A wild wheat harvest in Turkey. *Archaeology, 20*, 197–201.

Haslam, M. (2004). The decomposition of starch grains in soils: Implications for archaeological residue analyses. *Journal of Archaeological Science, 31*, 1715–1734.

Hastorf, C. A. (2017). *The social archaeology of food: Thinking about eating from prehistory to the present*. Cambridge: Cambridge University Press.

Hather, J. G. (1991). The identification of charred archaeological remains of vegetative parenchymous tissue. *Journal of Archaeological Science, 18*, 661–675.

Hather, J. G. (2016). *Archaeological parenchyma*. London: Routledge.

Havinga, A. J. (1984). A 20-year experimental investigation into the differential corrosion susceptibility of pollen and spores in various soil types. *Pollen et Spores, 26*, 541–558.

Helmling, S., Olbrich, A., Heinz, I., & Koch, G. (2018). Atlas of vessel elements: Identification of Asian timbers. *IAWA Journal, 39*(3), 249–352.

Henry, A. G. (2014). Formation and taphonomic processes affecting starch granules. In J. M. Marston, J. d'A. Guedes, & C. Warinner (Eds.), *Method and theory in paleoethnobotany* (pp. 35–50). Boulder, CO: University Press of Colorado.

Hevly, R. H. (1981). Pollen production, transport and preservation: Potentials and limitations in archaeological palynology. *Journal of Ethnobiology, 1*(1), 39–54.

Hillman, G. (1984). Interpretation of archaeological plant remains: Ethnographic models from Turkey. In W. Van Zeist & W. A. Casparie (Eds.), *Plants and ancient man: Studies in paleoethnobotany* (pp. 1–41). Rotterdam: Balkema.

Hillman, G. (2000). The plant food economy of Abu Hureyra 1 and 2. In A. M. T. Moore, G. C. Hillman, & A. J. Legge (Eds.), *Village on the Euphrates: From foraging to farming at Abu Hureyra* (pp. 327–329). Oxford: Oxford University Press.

Hoadley, R. B. (1990). *Identifying wood: Accurate results with simple tools*. Newtown: Taunton Press.

Höhn, A., & Neumann, K. (2018). Charcoal identification in a species-rich environment: The example of Dibamba, Cameroon. *IAWA Journal, 39*(1), 87–S47.

Horrocks, M. (2005). A combined procedure for recovering phytoliths and starch residues from soils, sedimentary deposits and similar materials. *Journal of Archaeological Science, 32*, 1169–1175.

Hosch, S., & Zibulski, P. (2003). The influence of inconsistent wet-sieving procedures on the macroremain concentration in waterlogged sediments. *Journal of Archaeological Science, 30*, 849–857.

Hubbard, E. M. (2010). Livestock and people in a middle chalcolithic settlement: A micromorphological investigation from Tel Tsaf, Israel. *Antiquity, 84*(326), 1123–1134.

Hubbard, R. N. L. B., & al-Azm, A. (1990). Quantifying preservation and distortion in carbonized seeds and investigating the history of friké production. *Journal of Archeological Science, 17*, 103–106.

Huffard, M. B., Xun, X., van Heerwaarden, J., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nature Genetics, 44*, 808–811.

Ilie, J. (1991). *CSIRO atlas of hardwoods*. Melbourne: Commonwealth Scientific and Industrial Research Organisation.

Jansen, D., & Nelle, O. (2014). The Neolithic woodland—Archaeoanthracology of six Funnel Beaker sites in the lowlands of Germany. *Journal of Archaeological Science, 51*, 154–163.

Kadane, J. B. (1988). Possible statistical contributions to paleoethnobotany. In C. A. Hastorf & V. S. Popper (Eds.), *Current paleoethnobotany: Analytical methods and cultural interpretations of archaeological plant remains* (pp. 206–214). Chicago: University of Chicago Press.

Katzenberg, M. A., Schwarcz, H. P., Knyf, M., & Melbye, F. J. (1995). Stable isotope evidence for maize horticulture and paleodiet in southern Ontario, Canada. *American Antiquity, 60*(2), 335–350.

Kislev, M. E., Hartmann, A., & Bar-Yosef, O. (2006). Early domesticated fig in the Jordan Valley. *Science, 312*, 1372–1374.

Kovach, M. J., Sweeney, M. T., & McCouch, S. R. (2007). New insights into the history of rice domestication. *Trends in Genetics, 23*(11), 578–587.

Lake, M. (2015). *Australian rainforest woods. Characteristics, uses and their identification*. Melbourne: Commonwealth Scientific and Industrial Research Organisation.

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology, 30*, 191–202.

Lamb, J., & Loy, T. (2005). Seeing red: The use of Congo red dye to identify cooked and damaged starch grains in archaeological residues. *Journal of Archaeological Science, 32*, 1433–1440.

Lawrence, R. J. (2003). Human ecology and its applications. *Landscape and Urban Planning, 65*, 31–40.

Lee, G. A. (2012). Taphonomy and sample size estimation in paleoethnobotany. *Journal of Archaeological Science, 39*, 648–655.

Lewis, M. C., & Sealy, J. C. (2018). Coastal complexity: Ancient human diets inferred from Bayesian stable isotope mixing models and a primate analogue. *PLoS ONE*, *13*(12), e0209411.

Lewontin, R. C. (1983). Gene, organism and environment. In D. S. Bendall (Ed.), *Evolution from molecules to men* (pp. 273–286). Cambridge: Cambridge University Press.

Lubbock, J. (1865). *Prehistoric times, as illustrated by ancient remains, and the manners and customs of modern savages*. London: Williams and Norgate.

Madella, M., Lancelotti, C., & Savard, M. (Eds.). (2014). *Ancient plants and people: Contemporary trends in archaeobotany*. Tucson: University of Arizona Press.

Margaritis, E. (2013). Distinguishing exploitation, domestication, cultivation and production: The olive in the third millennium Aegean. *Antiquity, 87*(337), 746–757.
Marguerie, D. (2002). Fuel from protohistorical and historical kilns in North-Western France. In J. S. Thiébault (Ed.), *Charcoal analysis: Methodological approaches, palaeoecological results and wood uses* (BAR international series 1063). Oxford: Archaeopress.
Marston, J. M. (2014). Ratios and simple statistics in paleoethnobotanical analysis: Data exploration and hypothesis testing. In J. M. Marston, d'A. Guedes, & C. Warinner (Eds.), *Method and theory in paleoethnobotany* (pp. 163–180). Boulder, CO: University Press of Colorado.
Matthews, B., De Meester, L., Jones, C. G., Ibelings, B. W., Bouma, T. J., Nuutinen, V., et al. (2014). Under niche construction: An operational bridge between ecology, evolution and ecosystem science. *Ecological Monographs, 84*(2), 245–263.
McGovern, P., Jalabadze, M., Batiuk, S., Calahan, M. P., Smith, K. E., Hall, G. R., et al. (2017). Early Neolithic wine of Georgia in the South Caucasus. *Proceedings of the National Academy of Sciences, 114*(48), E10309–E10318.
Metcalfe, D., & Heath, K. M. (1990). Microrefuse and site structure: The hearths and floors of the Heartbreak Hotel. *American Antiquity, 55*(4), 781–796.
Milla, R., Osborne, C. P., Turcotte, M. M., & Violle, C. (2015). Plant domestication through an ecological lens. *Trends in Ecology & Evolution, 30*(8), 463–469.
Miller, N. F. (1988). Ratio in paleoethnobotanical analysis. In C. A. Hastorf & V. S. Popper (Eds.), *Current Paleoethnobotany: Analytical methods and cultural interpretations of archaeological plant remains* (pp. 72–85). Chicago: University of Chicago Press.
Minnis, P. E. (1989). Prehistoric diet in the northern southwest: Macroplant remains from four corners feces. *American Antiquity, 54*(3), 543–563.
Moore, J. (2001). Can't see the wood for the trees: Interpreting woodland fire history from microscopic charcoal. In U. Albarella (Ed.), *Environmental archaeology: Meaning and purpose* (pp. 211–228). New York: Springer.
Moore, P. D., Webb, J., & Collinson, M. C. (1991). *Pollen analysis: A laboratory manual*. Oxford: Blackwell.
Morehart, C. T., & Morell-Hart, S. (2015). Beyond the ecofact: Toward a social paleoethnobotany in Mesoamerica. *Journal of Archaeological Method and Theory, 22*(2), 483–511.
Nesbitt, M. (2009). *Identification guide for Near Eastern grass seeds*. London: UCL Institute of Archaeology.
Odling-Smee, J. (1988). Niche constructing phenotypes. In H. C. Plotkin (Ed.), *The role of behavior in evolution* (pp. 73–132). Cambridge: MIT Press.
Odling-Smee, J., Laland, K., & Feldman, M. (2003). *Niche construction: The neglected process in evolution*. Princeton: Princeton University Press.
Orton, C. (2000). *Sampling in archaeology*. Cambridge: Cambridge University Press.
Pagán-Jiménez, J. R., Rodríguez-Ramos, R., Reid, B. A., van den Bel, M., & Hofman, C. L. (2015). Early dispersals of maize and other food plants into the southern Caribbean and northeastern South America. *Quaternary Science Reviews, 123*, 231–246.
Pearsall, D. M. (2014). Formation processes of pollen and phytoliths. In J. M. Marston, J. d'A. Guedes, & C. Warinner (Eds.), *Method and theory in paleoethnobotany* (pp. 51–73). Boulder, CO: University Press of Colorado.
Pearsall, D. M. (2015). *Paleoethnobotany: A handbook of procedures* (3rd ed.). Walnut Creek, CA: Left Coast Press.
Pearsall, D. M. (2019). *Case studies in paleoethnobotany: Understanding ancient lifeways through the study of phytoliths, starch, macroremains, and pollen*. New York: Routledge.
Peleg, Z., Fahima, T., Korol, A. B., Abbo, S., & Saranga, Y. (2011). Genetic analysis of wheat domestication and evolution under domestication. *Journal of Experimental Botany, 62*(14), 5051–5061.
Piperno, D. R. (2006). *Phytoliths: A comprehensive guide for archaeologists and paleoecologists*. New York: AltaMira Press.
Pourkheirandish, M., Hensel, G., Kilian, B., Senthil, N., et al. (2015). Evolution of the grain dispersal system in barley. *Cell, 162*(3), 527–539.
Punugganan, M. D., & Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature, 457*(7231), 843–848.
Ranere, A. J., Piperno, D. R., Holst, I., Dickau, R., & Iriarte, J. (2009). The cultural and chronological context of early Holocene maize and squash domestication in the central Balsas River valley, Mexico. *Proceedings of the National Academy of Sciences, 106*(13), 5014–5018.
Reichert, E. T. (1913). *The differentiation and specificity of starches in relation to genera, species, etc*. Washington: Carnegie Institution of Washington.
Rindos, D. (1984). *The origins of agriculture: An evolutionary perspective*. Orlando, FL: Academic.
Ruffinatto, F., & Crivellaro, A. (2019). *Atlas of macroscopic wood identification, with a special focus on timbers used in Europe and CITES-listed species*. Springer.
Sang, T., & Ge, S. (2007). The puzzle of rice domestication. *Journal of Integrative Plant Biology, 49*(6), 760–768.
Scheel-Ybert, R. (2016). Charcoal collections of the world. *IAWA Journal, 37*(3), 489–505.
Schopmeyer, C. S. (1974). *Seeds of woody plants in the United States* (Agriculture handbook no. 450). Washington: Forest Service, U.S. Department of Agriculture.
Schweingruber, F. H. (1990). *Microscopic wood anatomy. Structural variability of stems and twigs in recent and subfossil woods from central Europe* (3rd ed.). Birmensdorf, CH: Swiss Federal Institute for Forest, Snow and Landscape Research.
Shevchenko, A., Yang, Y., Knaust, A., Verbavatz, J. M., Mai, H., Wang, B., et al. (2017). Open sesame: Identification of sesame oil and oil soot ink in organic deposits of Tang dynasty lamps from Astana necropolis in China. *PLoS One, 12*(2), e0158636.
Sluyter, A., & Dominguez, G. (2006). Early maize (*Zea mays* L.) cultivation in Mexico: Dating sedimentary pollen records and its implications. *Proceedings of the National Academy of Sciences, 103*(4), 1147–1151.
Smart, T. L., & Hoffman, E. S. (1988). Environmental interpretation of archaeological charcoal. In C. A. Hastorf & V. S. Popper (Eds.), *Current palaeoethnobotany: Analytical methods and cultural interpretations of archaeological plant remains* (pp. 167–205). Chicago: University of Chicago Press.
Smith, W. (2001). When method meets theory: The use and misuse of cereal producer/consumer models in archaeobotany. In U. Albarella (Ed.), *Environmental archaeology: Meaning and purpose* (pp. 283–298). Dordrecth: Springer.
Smith, B. (2012). A cultural niche construction theory of initial domestication. *Biological Theory Online, 6*, 260–271.
Smith, B. (2014). Documenting human niche construction in the archaeological record. In J. M. Marston (Ed.), *Method and theory in paleoethnobotany* (pp. 355–370). Boulder, CO: University Press of Colorado.
Smith, P. (2018). *The book of seeds: A life-size guide to six hundred species from around the world*. Chicago: University of Chicago Press.
Sonnante, G., Hammer, K., & Pignone, D. (2009). From the cradle of agriculture a handful of lentils: History of domestication. *Rendiconti Lincei, 20*(1), 21–37.
Spooner, D. M., McLean, K., Ramsay, G., Waugh, R., & Bryan, G. J. (2005). A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. *Proceedings of the National Academy of Sciences, 102*(41), 14694–14699.

Stuijts, I. (2006). Charcoal sampling sites and procedures: Practical themes from Ireland. In A. Dufraisse (Ed.), *Charcoal analysis: New analytical tools and methods for archaeology* (BAR international series 1483) (pp. 25–33). Oxford: Archaeopress.

Szabó, P. (2015). Historical ecology: Past, present and future. *Biological Reviews, 90*(4), 997–1014.

Théry-Parisot, I., Chabal, L., & Chrzavzez, J. (2010). Anthracology and taphonomy, from wood gathering to charcoal analysis. A review of the taphonomic processes modifying charcoal assemblages, in archaeological contexts. *Palaeogeography, Palaeoclimatology, Palaeoecology, 291*(1–2), 142–153.

Torrence, R., & Barton, H. (Eds.). (2006). *Ancient starch research*. Walnut Creek: Left Coast Press.

Torrence, R., Wright, R., & Conway, R. (2004). Identification of starch granules using image analysis and multivariate techniques. *Journal of Archaeological Science, 31*, 519–532.

Ucchesu, M., Orrù, M., Grillo, O., Venora, G., Paglietti, G., Ardu, A., et al. (2016). Predictive method for correct identification of archaeological charred grape seeds: Support for advances in knowledge of grape domestication process. *PLoS One, 11*(2), e0149814.

Van der Veen, M., & Fieller, N. (1982). Sampling seeds. *Journal of Archaeological Science, 9*, 287–298.

VanDerwarker, A. M., Bardolplh, D. N., Hoppa, K., Thakar, H., Martin, L., Jaqua, A. et al. (2016). New World paleoethnobotany in the new millennium (2000-2013). *Journal of Archaeological Research, 24*(2), 125–177.

Vayda, A. P. (1969). An ecological approach in cultural anthropology. *Bucknell Review, 17*(1), 112–119.

Wang, H., Nussbaum-Wagler, T., Li, B., Zhao, Q., Vigouroux, Y., & Faller, M. (2005). The origin of the naked grains of maize. *Nature, 436*(7051), 714–719.

Watson, P.-J. (1976). In pursuit of prehistoric subsistence: A comparative account of some contemporary flotation techniques. *Midcontinental Journal of Archaeology, 1*(1), 77–100.

Weber, R. W. (1998). Pollen identification. *Annals of Allergy, Asthma & Immunology, 80*(2), 141–148.

Weisskopf, A. (2017). A wet and dry story: Distinguishing rice and millet arable systems using phytoliths. *Vegetation History and Archaeobotany, 26*(1), 99–109.

Wheeler, E. A., & Baas, P. (1998). Wood identification—A review. *IAWA Journal, 19*(3), 241–264.

Willcox, G. (2012). Searching for the origins of arable weeds in the Near East. *Vegetation History and Archaeobotany, 21*(2), 163–167.

Willcox, G. (2013). The roots of cultivation in southwestern Asia. *Science, 341*(6141), 39–40.

Willcox, G., Fornite, S., & Herveux, L. (2008). Early Holocene cultivation before domestication in northern Syria. *Vegetation History and Archaeobotany, 17*, 313–325.

Winterhalder, B., & Smith, E. A. (2000). Analyzing adaptive strategies: Human behavioral ecology at twenty-five. *Evolutionary Anthropology, 9*(2), 51–72.

Wright, P. J. (2005). Flotation samples and some paleoethnobotanical implications. *Journal of Archaeological Science, 32*(1), 19–26.

Zheng, Y., Crawford, G. W., Jiang, L., & Chen, X. (2016). Rice domestication revealed by reduced shattering of archaeological rice from the lower Yangtze Valley. *Scientific Reports, 6*, 28136.

17 Soils, Sediments, and Geoarchaeology

> *… worms have played a considerable part in the burial and concealment of several Roman and other old buildings in England; but no doubt the washing down of soil from the neighbouring higher lands, and the deposition of dust, have together aided largely in the work of concealment*
>
> (Darwin 1882: 230).

The soil and sediment that archaeologists excavate are not just a passive, artifact-bearing medium to sift and ignore; as even Charles Darwin understood, they are deeply implicated in site-formation processes, paleoenvironment, human activities, artifact conservation, and chronology. Some aspects of geology and geomorphology are so essential to modern archaeology that they spawned the subdiscipline of geoarchaeology (see Butzer 2008; Courty et al. 1989; Fouache 2013; French 2003; Goldberg and Macphail 2006; Goldberg et al. 1993; Holliday 2004; Rapp and Hill 2006; Stein and Farrand 2001; Waters 1992). Notably, geoarchaeology is not at all restricted to site contexts; it is also essential for understanding ancient landscapes and sites' places in them. While most of this is the work of geoarchaeological specialists, it is important for archaeologists to understand the fundamentals as they apply to both fieldwork and laboratory work.

17.1 Soils and Sediments

While many people use these terms interchangeably, "soil" and "sediment" are not synonyms. Soils are the product of *in situ* weathering of the Earth's crust, while sediments are deposits of particles that various processes have removed from other locations and caused to accumulate elsewhere (Shackley 1975; Holliday 2004). However, after sufficient time, sediments can themselves undergo weathering processes that cause them to turn into soils. Archaeological, artifact-bearing deposits, for example, sometimes acquire the characteristics of soils after thousands of years. Geologists consider the sedimentary cycle to consist of weathering of the Earth's crust, and transport, deposition, and post-depositional alteration of particles.

Weathering processes lead to zonation, one of the key characteristics of soils, with zones called "horizons" in vertical profiles. The **A horizon** at the top of the profile is richest in humus, but rainwater filtering through this horizon has removed some of its components through **leaching**. The **B horizon** is the zone in which these leached components, especially certain minerals and soil nutrients, are redeposited. The **C horizon** is the zone where we find weathered bedrock, just above the bedrock that is parent material for the soil. It is chemically the same as the bedrock because no leaching has either removed or redeposited chemicals there. Where there are thinner zones within these three horizons, convention calls for us to designate them as A_1, A_2, B_1, B_2, etc. A soil **profile** describes the internal structure of a soil, not a sequence of stratified layers or deposits, which instead should be called a **stratigraphic series** (Shackley 1975: 4; and Chap. 19).

17.2 Lithostratigraphy and Archaeological Stratigraphy

Geologists call the subdivisions of the Earth's crust, as defined by their physical, lithologic characteristics, lithostratigraphic units (NACSN 1983; Stein 1987). Soil scientists instead distinguish pedological units that consist of mineral and organic materials that climate and living organisms have altered through soil-forming processes (Fig. 17.1). In a soil profile, the horizons are different pedological units, whether they formed in place, were deposited at the same time or different times, or gradually accumulated over time.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_17

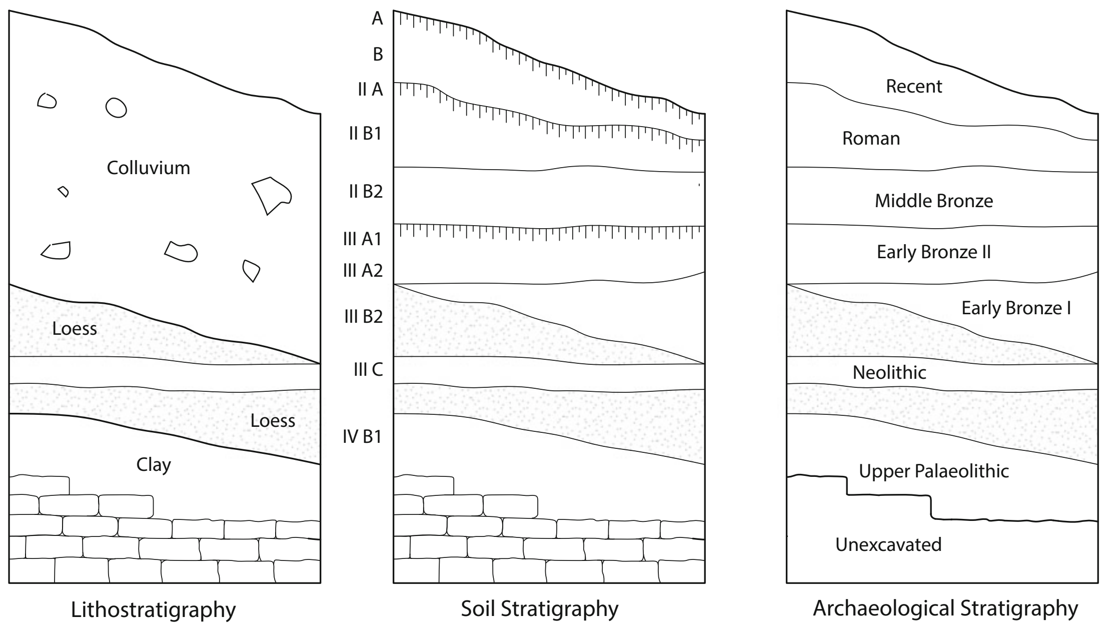


Fig. 17.1 Comparison of lithological, pedological and archaeological units in a profile or stratigraphic series. (Modified from Courty et al. 1989)

Although not all authors agree that pedological units are different from lithostratigraphic ones, soil scientists routinely subdivide soil profiles more finely than a geologist would, and pedological units do not necessarily represent distinct depositional events. Geologists also talk about biostratigraphic units, defined by their fossil content.

Archaeologists, meanwhile, routinely subdivide profiles or stratigraphic series even more finely than soil scientists do, and rely not only on lithological characteristics, such as color and texture, but also artifact content, much as geologists consider fossils. Stein (1990), prefers to call these archaeological units "ethnostratigraphic units," but other terms for them include anthropogenic units (formed by human agency), anthropic soils (soils in which human action is a major soil-forming factor), and archaeological or cultural layers, strata, loci, and features. Archaeologists often define these last units by a combination of lithological, pedological and cultural indicators (see Chap. 19). Typically, field archaeologists privilege such characteristics as color, texture, and particle sorting for defining stratigraphic units, and then consider artifact content afterwards.

17.3 Sampling Sediments for Laboratory Analysis

Although geoarchaeologists make many of their observations in the field, some kinds of analyses take place in the lab, and require fieldworkers to remove volumes of sediment for these analyses.

The most common type of sample consists of volumes of sediment removed either from auger holes or during the course of either archaeological excavation or a geoarchaeologist's sampling of an exposed profile. **Bulk samples** generally consist of a number of bags of loose sediment, each ideally taken from within a particular stratigraphic unit, that have lost any evidence for the original arrangement of their particles (see also p. 272). Augering, for example, churns up the particles by its rotary action but preserves the particles themselves, along with their chemical characteristics. Removing material with trowel or shovel also disturbs sediments and scrambles their physical arrangements. Bulk sampling is useful for a wide range of purposes, including examination of texture, particle shape, and chemical composition.

Core sampling, by contrast, involves removing long, cylindrical volumes of soil and sediment in metal tubes that pierce the ground vertically. Because insertion of the coring tube disturbs the sediments much less than augering does, core sampling yields segments that largely preserve the arrangement of particles and interfaces between layers. However, because most cores have small diameters, they do not work well in soils and sediments that contain a lot of stones larger than the core diameter.

Another way to sample soils and sediments without losing information about their structure is to use **block sampling**. This involves removing blocks of material without disturbing their arrangement, wrapping them tightly to prevent them from falling apart, and later examining them in thin-section. Further description of this method occurs below (p. 302).

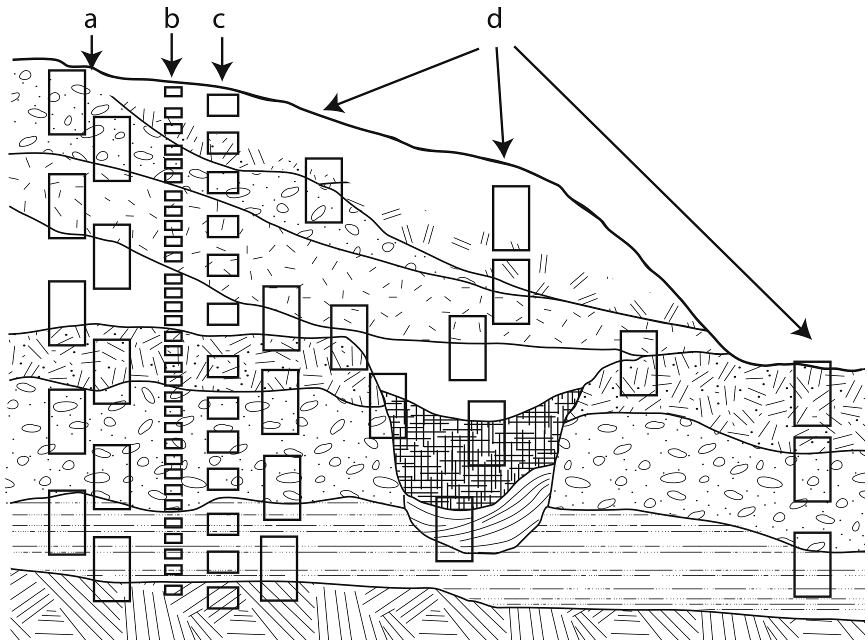


Fig. 17.2 A variety of sampling designs for removal of block or bulk samples from a stratified profile: (**a**) systematic sampling of blocks for micromorphology, (**b**) systematic bulk sampling for pollen analysis, (**c**) systematic bulk sampling for chemical or particle analysis, and (**d**) purposive sampling of blocks for micromorphology, emphasizing stratigraphic transitions. (Modified from Courty et al. 1989: Fig. 3.9)

As with any kind of sampling (see Chap. 6), it is important to have a sampling strategy appropriate to your research questions. Much geoarchaeology involves a form of purposive sampling that targets, for example, major changes in sedimentary processes below and above interfaces or the identification of unusual deposits and the origin of their particles. For example, it is often preferable to select locations for block sampling where the blocks will intersect transitions between layers. Many archaeologists use sediment samples to assist their interpretation of site stratigraphy, so stratigraphic questions can be a major factor in selecting sediments for further analysis. Different research questions could lead to a variety of possible purposive or systematic sample designs (Fig. 17.2).

17.4 Texture and Particle Characteristics

Archaeological sediments consist of accumulated particles that include tiny clay minerals, broken artifacts and animal bones, ash, charcoal fragments, pebbles and rocks that originate either from the weathering of underlying bedrock or from human and natural transport. **Clasts** are fragments that weathering processes have removed from parent rock and, in a broad sense, include most artifacts as well as non-cultural pebbles and cobbles. There can also be voids that result from the action of plants and burrowing animals.

17.4.1 Texture

One important characteristic of sediments is their texture. This is a statement about the particle-size distribution in the sediment (Inman 1962), which we can determine by a detailed particle-size analysis or estimate less precisely by wetting, squeezing and working a small ball of the sediment in our hands to characterize it on a nominal scale (Fig. 17.3). Some of the major categories on such a scale are as follows.

Sand has particle sizes in the range of 4–0.5 ϕ, and the lack of smaller particles to hold it together makes it loose, especially when dry, so it is impossible to squeeze it into a ball that will not fall apart. If you rub sand on your skin, you can feel its grittiness, especially with sand in the range of 1–0.5 ϕ.

Silt has a smaller particle size than sand (8–4 ϕ), with grains that are not visible to the naked eye. Although it can be somewhat gritty, it generally feels smoother or silkier than sand when rubbed between the fingers.

Clay has a still smaller particle size, greater than 8 ϕ, which makes it sticky and plastic when wet. When dry, it forms hard lumps with shrinkage cracks.

Sandy Loam contains enough clay and silt to hold the sand together, with about 50% sand, 30% silt, and 20% clay (Shackley 1975: 12).

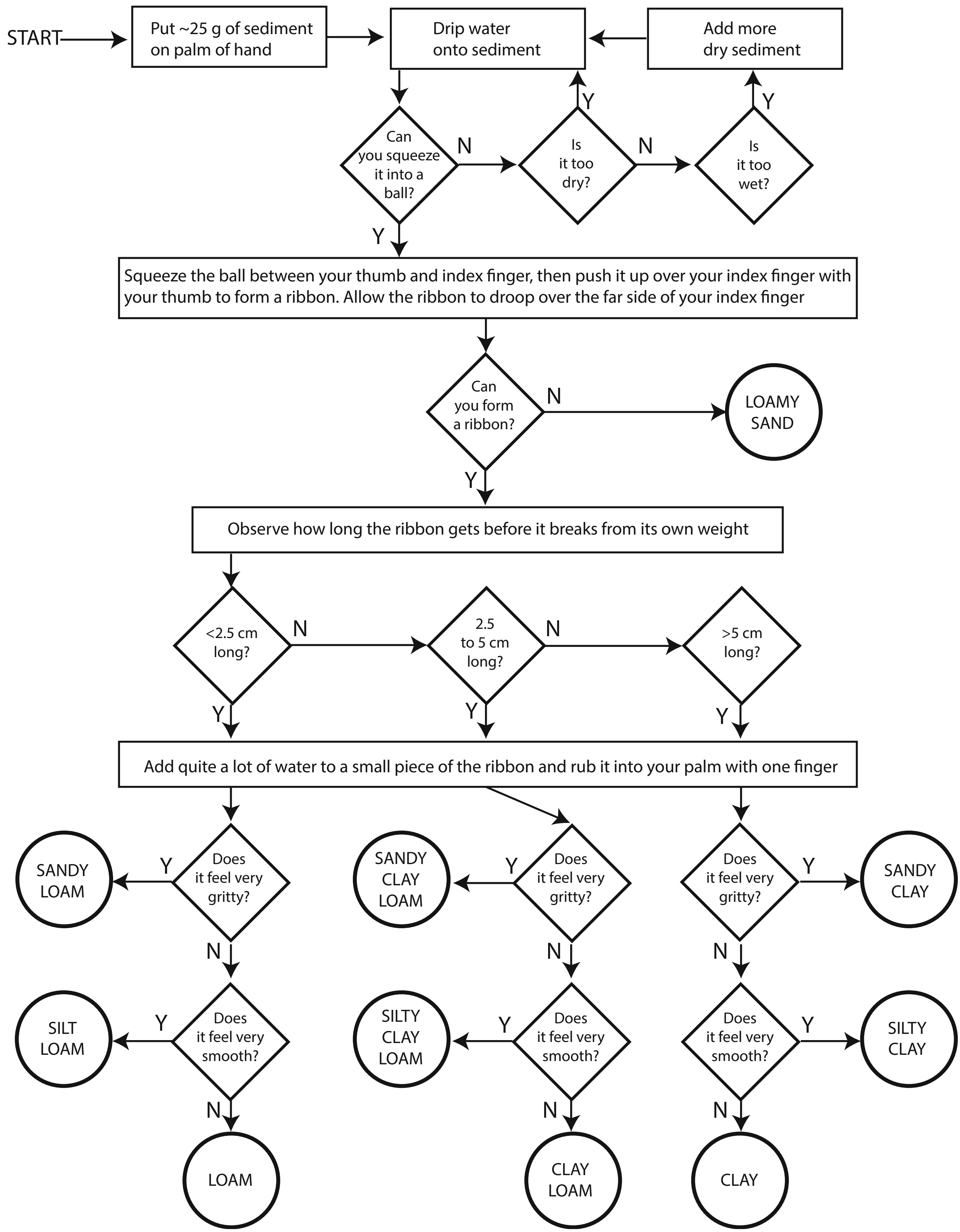


Fig. 17.3 Step-by-step program to help classify sediment textures. (Modified after Thien 1979)

Loam consists of roughly 40% sand, 40% silt and 20% clay, making it only somewhat gritty and more plastic than sandy loam.

Silty loam contains at least 50% silt and sand, with 12–25% clay that makes it feel silky when wet but form clods when dried out. The clods break into a floury powder.

Clay loam has roughly equal parts sand and clay, so that it is plastic and holds together when wet but forms hard clods when dry.

One of the reasons that texture is so important is that particle size helps us identify the agent, environment and process of sediment deposition (Allen 1968; Friedman 1961; Mason and Folk 1958). Another is that it helps archaeologists determine which deposit in one excavation area may correspond with one in another excavation area that may have been created as part of the same geological or cultural process.

We can define particle "size," however, in several different ways (Fig. 17.4; Folk 1966). Much as with size measures on artifacts, how we define such things as "diameter" or volume can result in noticeably different size distributions.

While modern geomorphological laboratories determine the grain-size distributions of soils and sediments with laser granulometers, which only require small sample elements and have few health-and-safety concerns, these are expensive and rarely found in archaeological laboratories. The most common grain-size measurement among archaeologists remains sieve diameter, the smallest screen aperture that allows the particle to pass through in a stack of nested screens, with the largest mesh on top and smallest at the bottom. This can be deceiving because particles are not often spherical, and smaller particles may also adhere to larger ones and get caught in sieves that they "should have" passed through. This method also poses the problem that it cannot separate silt from clay, requiring analysts to use hydrometer analysis of the finest material if making a quantitative silt/clay distinction is necessary. However, because it is relatively easy and inexpensive, screening remains a popular choice. Sieve sets come in a variety of standard aperture series: the ASTM (American Society for Testing Materials), Tyler, and British Standard, among others (Table 17.1). Usually they are brass cylinders about 20 cm in diameter, with a mesh of woven brass or steel or, for finer apertures, phosphor bronze or nickel mesh.

The decision on whether to use dry or wet sieving depends on the risk to things like carbonized seeds or tiny artifacts that the sediment might contain. Wet sieving is usually best for fine-grained sediments that tend to clump when dry. There are also health concerns around dry-screening silts and clays as it can cause clouds of fine dust that are a respiratory risk. Wear a dust mask. Sometimes archaeologists screen the "raw" sediment, but some kinds of sediments may call for preparation, such as removal of carbonates ($CaCO_3$; e.g., Stein 1990: 142).

To dry-sieve a volume of sediment, begin by measuring the mass of each clean, empty sieve and the collection pan on a properly calibrated electronic balance, and then stack the pan and sieves on the screen shaker in serial order from finest, at the bottom, right above the pan, to coarsest, on top. If you are using a large number of sieves, you may need to break the job down into subsets, beginning with the coarsest series. Try to ensure that the total volume you will sieve will be enough to get a decent sample size on each screen, and dry the sediment in a lab oven set at 105 °C for several hours, or leave it in clean, cloth bags in a warm, dry room for several days. If your sediment contains a lot of damp clay, you may

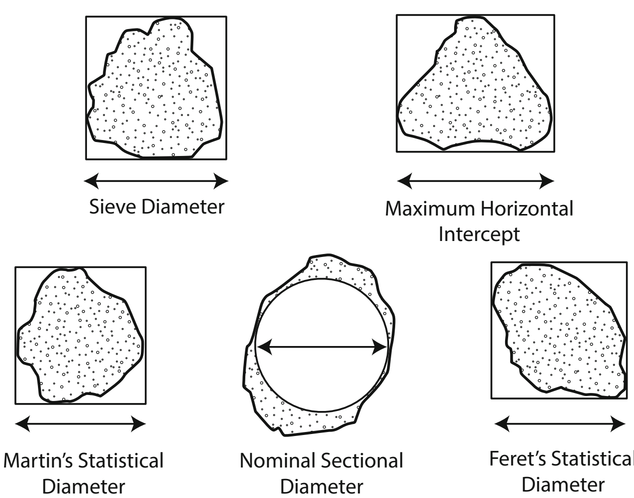


Fig. 17.4 Different ways to measure particle size. (Y. Salama, after Shackley 1975: 89)

Table 17.1 U.S.Standard sieve sizes

US sieve size	mm	US sieve size	mm
4	4.76	45	0.354
5	4.00	50	0.297
6	3.36	60	0.250
7	2.83	70	0.210
8	2.38	80	0.177
10	2.00	100	0.149
12	1.68	120	0.125
14	1.41	140	0.105
16	1.19	170	0.088
18	1.00	200	0.074
20	0.841	230	0.063
25	0.707	270	0.053
30	0.595	325	0.044
35	0.500	400	0.037
40	0.420		

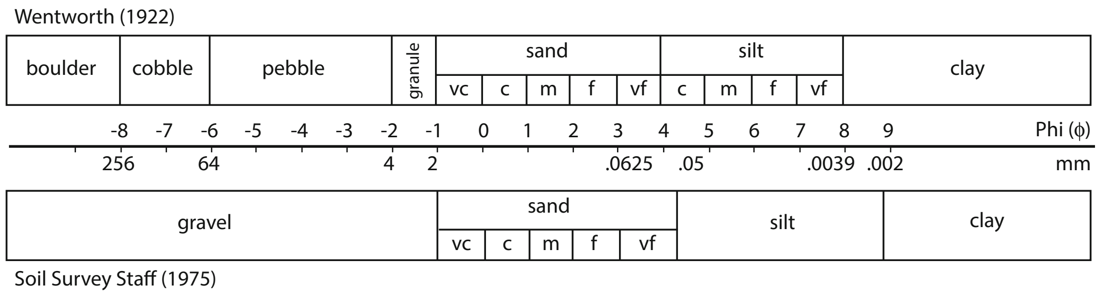


Fig. 17.5 Competing scales for the measurement of particle size (after Courty et al. 1989: 36). Abbreviations include vc (very course), c (coarse), m (medium), f (fine), vf (very fine), and ϕ (phi)

instead have to dry it in an oven set about 50 °C for several days, to prevent it from baking into a brick.

Pour the sediment volume into the uppermost, coarsest sieve, cover it with a lid, and secure the stack of screens with clamps. Set a timer for 10 min and turn the screen shaker on. After 10 min, remove the sieves and remeasure the mass of each, now including a portion of the sediment. Record the net mass (gross mass - mass of empty screen) for each screen, remembering that it contains (ideally) all the particles bigger than that screen aperture, but smaller than the aperture of the sieve above it.

Then, carefully empty the sieves into labelled sample bags (assuming you want to examine them for micro-artifacts, charred plant remains, fish vertebrae, etc.), brush the sieves gently to remove any adhering clay particles, and they are ready for your next volume.

Although the size distributions in a sediment are continuous, it is common to express them on an ordinal scale (Fig. 17.5), such as the Wentworth (1922a, b), British Standard (BSI 1999), European Standard, Soil Survey Staff (1975), or Unified Soil Classification (ASTM 2011) scale. Shackley (1975: 91–92) prefers the logarithmic Krumbein (1934) scale (units of ϕ).

If you choose to display the ratio-scale data on particle size in a histogram, remember that the sieve series has unequal intervals. Consequently, ensure that you represent the quantities by area, either with a key showing how much area corresponds to a particular mass (Fig. 17.6), or by scaling the heights of bars, on the y-axis, with ratio units such as grams per mm.

Alternatively, you could use a line graph that was explicitly designed with particle sizes in mind. These are cumulative frequency graphs, usually with a logarithmic x-axis better to accommodate both very large (e.g., cobbles) and very small (clay) particles (Fig. 17.7).

Finally, yet another common way to display sediment texture is the ternary graph (Fig. 17.8), which is particularly good for showing how the textures of different sediments compare and whether or not they form clusters.

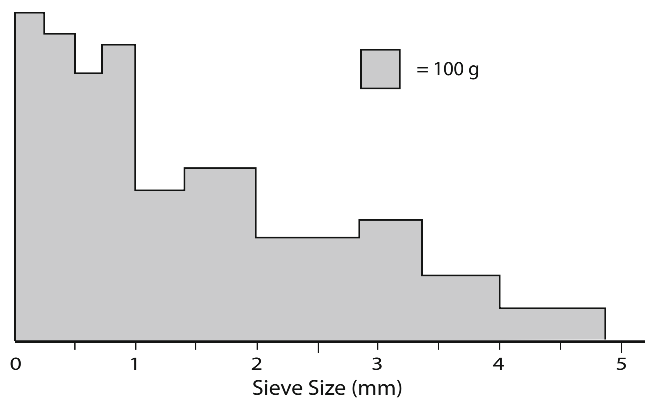


Fig. 17.6 Histogram of particle sizes in a sediment volume. Note that the x-axis has unequal intervals determined by the sieves used, but the square above the graph is a key to indicate how much area in the histogram corresponds with 100 g of particles

Control over Measurement Error

To ensure the accuracy of the mesh, sieves need care; using the same sieves for dry and wet sieving, for example, is inadvisable, as corrosion on the mesh will tend to narrow the apertures, leading to bias. Careless handling, such as rubbing or prodding particles on the mesh, or loading the mesh with too much sediment at once, can stretch or tear the mesh, also leading to bias. To check the sieves' accuracy, you can pass standard glass beads through the sieves or measure sampled areas of the mesh under magnification with a micrometer. Sieving sediment samples too long, especially with a mechanical shaker, can also cause bias by eroding or breaking the larger particles and by eroding the mesh itself. Run the screen shaker only for about 10 minutes (Shackley 1975: 109–111).

Alternatives such as maximum particle length, mean diameter and diameter of the largest sphere that could fit within the particle (Fig. 17.5) are not as simple to implement, requiring such methods as elutriation, a sorting of particles in streams of gas or liquid, for very small particles.

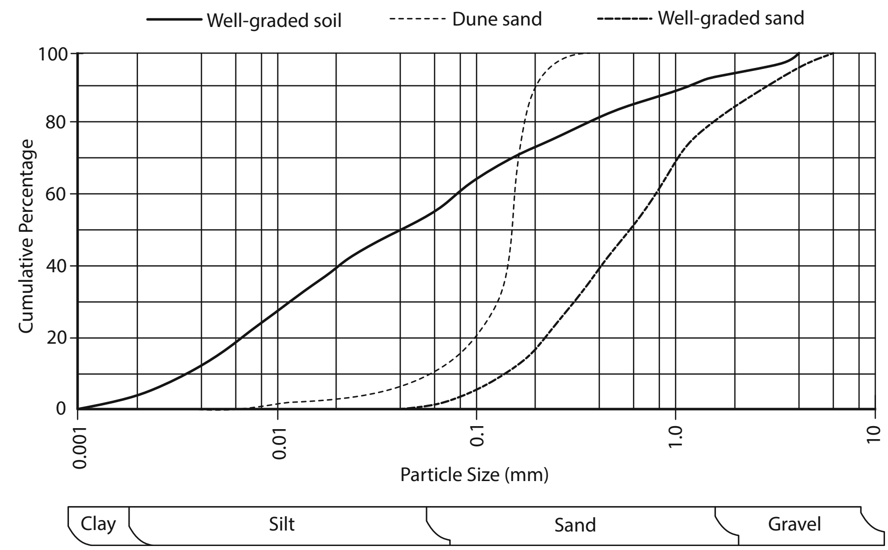


Fig. 17.7 Cumulative frequency graph of particle distributions of a well-graded soil, dune sand and well-graded sand. (Y. Salama)

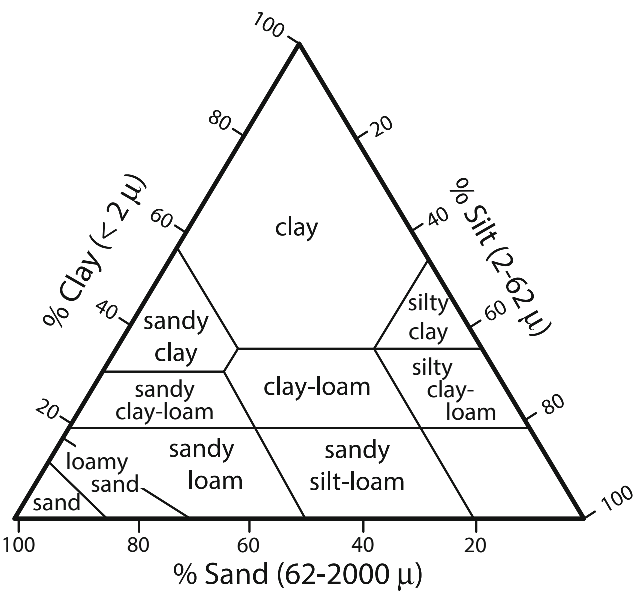


Fig. 17.8 Ternary graph to illustrate classes of sediment texture. (After Courty et al. 1989: 37)

17.4.2 Particle and Aggregate Shape

One simple way to characterize the shapes of larger particles, such as pebbles and cobbles, is by "aspect ratios" among three linear measurements: length along the long axis (l), maximum width perpendicular to length (w), and short axis (or thickness, s), taken with calipers or in a measuring frame. The ratios l/w and s/w give a very rough idea of shapes of individual stones, or you could characterize large numbers of particles with histograms, scatterplots or ternary graphs showing the three values. Such graphs can be useful to see whether pebbles from different archaeological contexts are likely to come from the same statistical population, and thus possibly from the same depositional event or site-formation process. A more complex alternative is to characterize mathematically the curvature of the particle outline (Durian et al. 2006), which effectively distinguishes degrees of rounding that could result from erosion of the particles.

Archaeologists more typically use ordinal scales for some of the attributes of particle shape, such as roundness (Fig. 17.9) and sphericity, using a visual chart (Krumbein 1941; Rittenhouse 1943). Roundness is a good clue to the depositional history and environment of a sediment (Shackley 1975: 46). Erosion causes wind-transported and water-transported particles to be more rounded than ones transported by humans or colluvial processes. Sphericity can also be useful to identify deposits transported by agents that eroded the particles.

In addition to visual guides, there are indices based on linear measurements, such as Krumbein's Intercept Sphericity (ψ_K), which is the cube root of the length (L), width (W), and short axis or thickness (T), divided by the square of the length:

$$\psi_K = \sqrt[3]{\frac{LWT}{L^2}}$$

Wadell (1932, 1933) proposed several indices, including "True" Sphericity, based on the ratio of the surface of a sphere of the same volume as the particle to the particle's actual surface area:

$$\psi = \frac{\sqrt[3]{36\pi V^2}}{S}$$

where V is the volume of the particle and S is its surface area. More complex analyses of roundness depend on cumulative

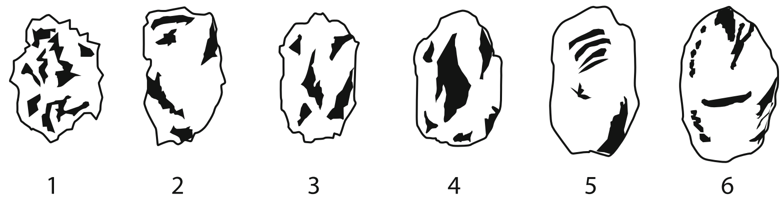


Fig. 17.9 Ordinal scale for estimating roundness of particles

curvature distributions, based on bounding boxes or the probability of finding curvatures within a particular range (Cruz-Matías et al. 2019; Durian et al. 2006).

Sometimes it is useful to employ a nominal scale to characterize the shapes of aggregates of particles (called **peds**, Fig. 17.10):

Platy: Plate-like particles with nearly horizontal grain surfaces.
Prismatic: Blocks with well-defined vertical faces (high length/width ratio) and angular vertices.
Columnar: Similar to prismatic, with strong vertical faces, but with more rounded vertices that give peds a pillar-like shape.
Angular blocky: Grain surfaces are fairly flat, vertices are angular, and both length/width and thickness/width ratios are close to 1.0, so the long axis is not obvious.
Subangular blocky: Similar to angular blocky, but with both flat and rounded surfaces and most of the vertices are rounded.
Granular: Particles are roughly spherical or polyhedral, and nonporous.
Crumbs: Aggregates of particles that are spherical or polyhedral, but porous.

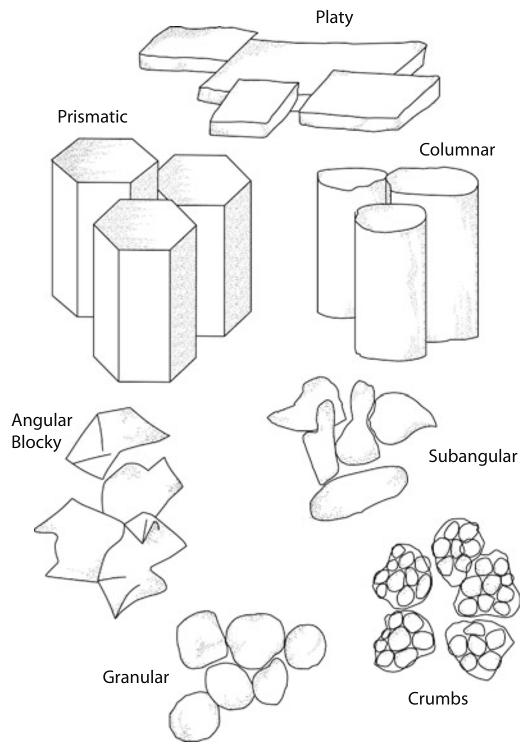


Fig. 17.10 Shape categories of particles and peds. (After Shackley 1975: 45)

17.4.3 Sorting and Density of Particles

It is also useful to estimate the density of pebbles or cobbles in a sediment, and their sorting. Well-sorted sediments are ones in which all the particles have a restricted size range (e.g., sand), while poorly sorted ones have a mixture of small, medium and large particles (Fig. 17.11). Sorting is a clue to the formation processes of a sediment. For example, alluvial sediments may be very well-sorted.

17.5 Sediment Color

Archaeologists commonly use a standard Munsell soil chart to measure the colors of sediments as well as pottery and lithics (Munsell Color 2010). The Munsell system characterizes colors by scales on three dimensions: hue, value (lightness or darkness) and chroma (departure from grey), making it a three-dimensional paradigmatic classification of color, but with ordinal dimensions (Fig. 17.12). To measure the color of a sediment sample, place a small amount of moist, but not wet, sediment on a clean trowel or spatula and hold it under the holes in the chart's pages (usually you can guess which page to start with) so that you can find the most closely matching color chip. Do this in natural, indirect sunlight unless you have special lamps meant for this purpose, and do not wear sunglasses. Record the three dimensions in the order of hue, value/chroma, e.g., 10YR 4/6.

Today, there are also hand-held digital meters that facilitate more consistent recording of Munsell colors. When used, as with the more traditional charts, under standardized conditions, these provide very reliable measurements of

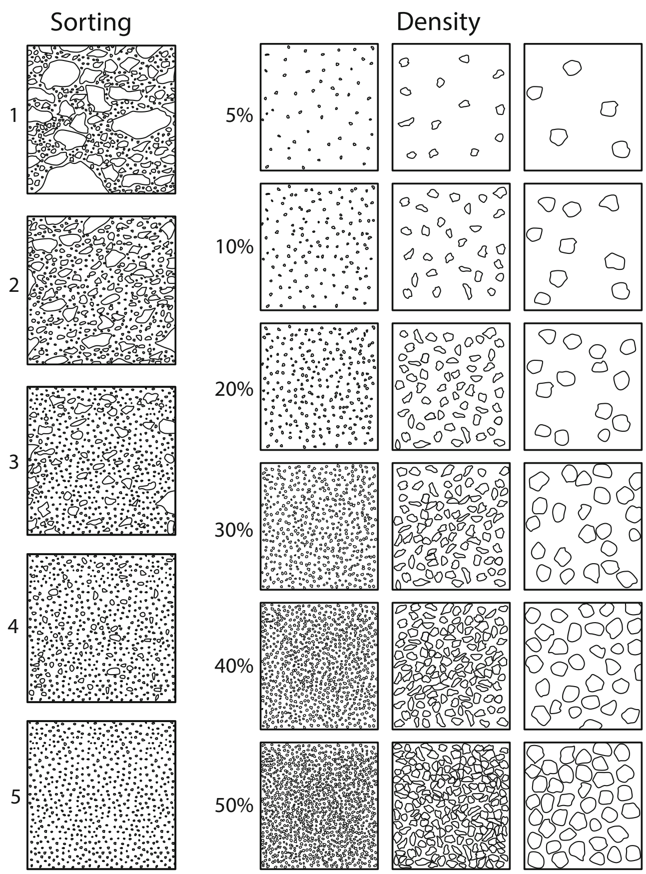


Fig. 17.11 Ordinal scales for estimating sorting and density of particles. (After Shackley 1975: 50)

color by analyzing the wavelengths of light reflected from sediments or other materials. Consequently, they can be an important element for ensuring the quality of color assessments. Even in a lab or field project that continues to use the traditional charts, it is possible to audit the chart-based measures periodically to ensure that assessments are within tolerances.

17.6 Acidity (pH)

pH is a measure of the concentration of hydrogen ions in the sediment, along a scale from 0 to 14 such that values less than 7 indicate acidity and ones higher than 7 indicate basic or alkali material. Most archaeological sediments have pH between 5 and 9 (Shackley 1975: 65). pH is one of the factors governing the probability that some materials, such as bone or pollen, will be preserved in the sediment. Consequently, it is important to know what pH your sediments have, especially if you suspect that **diagenesis**—the physical and chemical alteration of material in sediments—has affected the preservation of archaeologically important materials at your site.

There are several ways to measure pH, ranging in precision, accuracy, and ease of use. The simplest but least accurate and precise is to use litmus paper. After making a suspension of the sediment in distilled water, you dip the litmus paper in, shake off excess liquid, wait for the paper to change color, then compare this color with a chart. Typically, this allows easy measurement of pH in increments of 1 and with an error of $\pm$0.5 pH if you check your readings against a "buffer" solution of known pH value. However, even if you are careful, presence of proteins, salt, SO_2 and some other chemicals could result in biased measurements.

Somewhat better are indicator solutions. Some substances, such as phenol-phthalein, change color markedly when they reach different pH levels, but are difficult to use. A "universal" indicator, such as BDH, is easier to use. You mix a small amount of the sediment with some distilled water in a test tube and stir it into a suspension, then add a few drops of the suspension (avoiding large grains) to a piece of white, glazed tile. You then add a couple of drops of the indicator and compare the resulting color with standards (Shackley 1975: 66). Commercially available gardeners' test kits sometimes include indicator solutions.

However, the prices of digital pH meters have come down so much in recent years that it makes sense to use this much more reliable technology (Shackley 1975: 66–68). The meter measures the "effective" pH of the sediment, the electrical potential measured with a glass electrode, which is influenced by all sources of hydrogen ions. You calibrate the meter by inserting the electrode into standard "buffer" solutions of known pH, typically 4.0 and 9.0. After mixing equal parts of sediment and distilled water to make a slurry, you rinse the glass electrode with distilled water, insert it into the slurry, and read the pH value on the digital display. Never allow the electrode to dry out; meters come with a cap that you use to seal the end of the electrode after first putting a few drops of "storage liquid" or pH 7 buffer solution into it. If you are measuring pH on highly calcareous sediments, you may need to rinse the electrode with a mild acid, such as acetic acid (vinegar), once in a while to remove carbonate build-up on the electrode. If so, make sure you also rinse off the acid with distilled water.

17.7 Phosphates

Phosphate (or orthophosphate) is an ion of one phosphorus and four oxygen atoms (PO_4^{3-}) that derives from phosphoric acid. The phosphate content of soils and sediments is important because elevated phosphate concentrations result from decay of bone or feces, thus making them a clue to the presence of settlement sites, middens, and graves, even when diagenesis has caused bone to disappear entirely

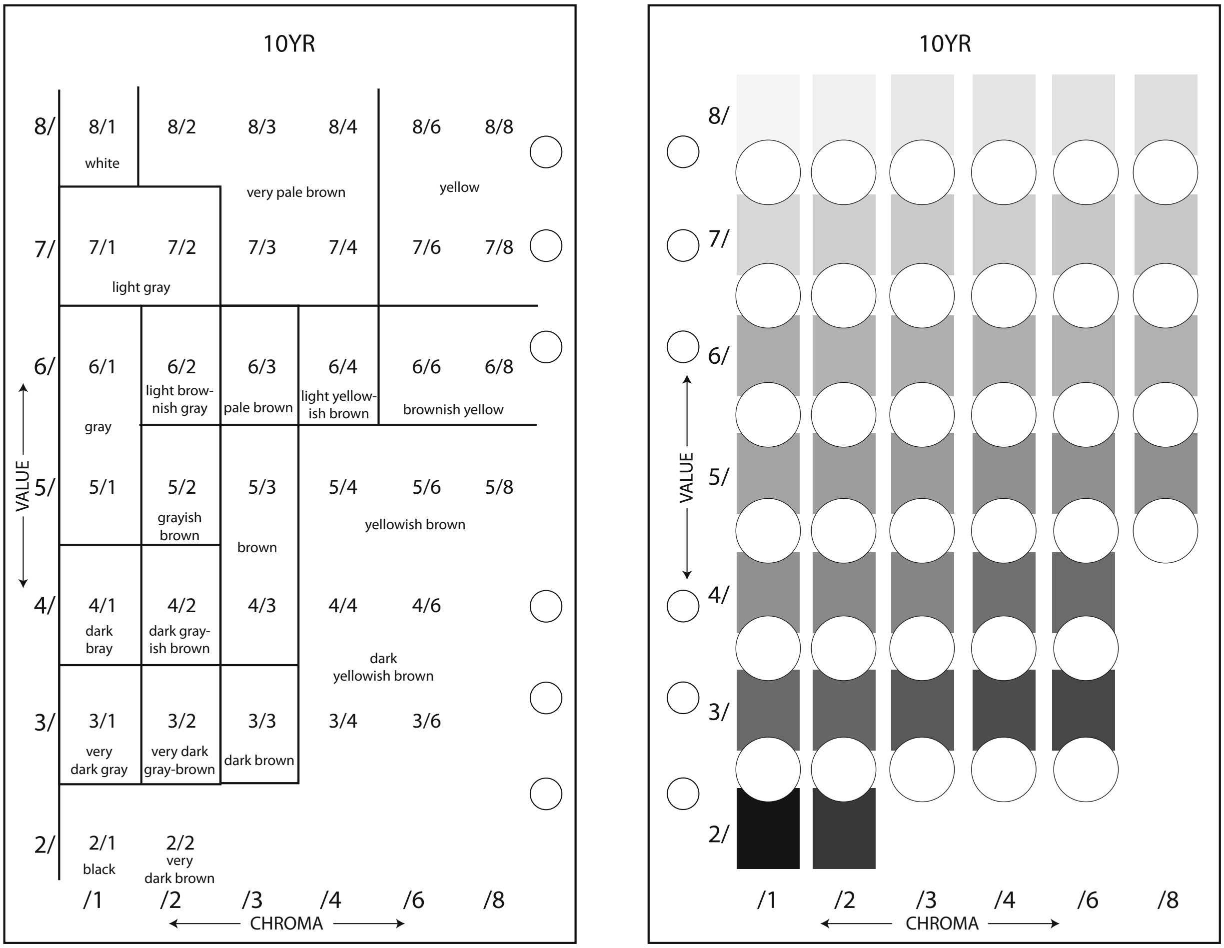


Fig. 17.12 Schematic of the 10YR page from a Munsell soil color chart, showing the arrangement of rows and columns for value and chroma. (After Munsell Color 2010)

(Davidson 1973). The "background" levels of phosphate in some soils (and the rocks from which they formed) are often quite low, so that archaeological sediments really stand out. However, some limestone-derived soils can have relatively high phosphate content as the result of fossil content, while it is also possible for phosphate to leach out or for modern fertilizers to contaminate soils and sediments with phosphate.

Measurement of phosphate involves converting the phosphorus compounds into dissolved orthophosphate by hydrolysis with acid and "digestion" through application of heat to break down the strong chemical bonds, and then doing colorimetry or chromatography. Boiling an extract of the sediment with ammonium molybdate and nitric acid (Shackley 1975: 68–70) is not an attractive or safe option for most archaeologists, but they may use the Gundlach test, which is suitable for phosphorus concentrations of 0.01 to 6 mg/L. It involves putting a small amount of sediment in a Petri dish and adding two drops of a solution of ammonium molybdate and sulphuric acid by pipette. After a couple of minutes, you add two drops of an ascorbic acid solution. As the liquids are absorbed into a filter paper, any phosphates in the sample will cause a color change to yellow and then blue. The amount of color change is a rough indication of the amount of total (organic and inorganic) phosphate present in the sample. Schwartz (1967; Shackley 1975: 78–69) presents a scale of five color-intensity increments. However, higher concentrations of phosphorus call for tests that use vanadomolybdophosphoric acid rather than ascorbic acid. Each method also varies in its precision (Rice et al. 2017: 241). Today, an excellent solution is to use a commercially available digital orthophosphate analyzer, which automates and simplifies these measurements.

17.8 Micromorphology

Today, archaeologists' interpretations of ancient deposits and site-formation processes, including human activities, depend a good deal on microscopic examination of profiles or stratigraphic series. To accomplish this, it is necessary to remove blocks of sediment that can be processed and made into thin-sections (Bullock et al. 1985; Courty et al. 1989; Goldberg 1992).

In the field, the blocks of soil or sediment can be removed intact in one of two ways. The first is to use a "Kubiena box," a metal box, open on one side, that one can hammer into a clean, vertical section at a strategic location and then carefully remove in such a way as to detach a mainly undisturbed block from the section. Kubiena boxes tend to work well when the soil or sediment is not too stony and the blocks are on the order of 10 cm in maximum dimension. The other method permits removal of somewhat larger blocks from well-consolidated sediment. It involves "channeling" the section by excavating two vertical trenches on either side of the area you would like to sample with a small pick or sharp knife, and then digging two shorter, horizontal ones above and below the desired block. Then you can break off the block and wrap it carefully with paper towel or cellophane secured with packing tape, or with moistened plaster cloth. With either method, you should mark the orientation of the box or block, usually by marking an "up" arrow, along with an identifying label or GPS coordinates.

Strategic locations for sampling micromorphology include certain or suspected interfaces between deposits (Fig. 17.2). To investigate a suspected house floor, for example, you would want a sediment block that included not only the floor itself, but the deposits above and below the floor. To understand thicker deposits or soil zones, you should take a series of overlapping blocks that cover most or all of their vertical extent.

Preparing the blocks removed by either method for thin-sectioning requires hardening it, generally by oven drying followed by impregnation with a polyester resin or epoxy under vacuum. Once it has achieved rock-like hardness, it is possible to cut a thin, flat slab with a rock saw, mount it on a glass slide, and polish it down to a thickness of 30 μm. The slide should also have the orientation marked on it.

Micromorphology has the advantage of great sensitivity in identification of the origin and depositional and post-depositional processes associated with soils and sediments (Goldberg 1992: 165). In archaeological contexts, it can, for example, help to identify house floors and some of the activities that occurred on them, reveal the presence of storage facilities, or distinguish mud bricks from mud and roofed from unroofed spaces.

17.9 Geomorphology and Site-Formation Processes

The cultural and natural processes that contribute to the formation of an archaeological site conform to some basic geological principles, and analysis of the resulting sediments helps us determine what those processes were.

Some of the processes that erode rock, soils and sediments are the same ones that transport the removed material to new locations: wind, water, gravity, and the activity of plants and animals, including humans. They can also act in concert; wind, rain and gravity can jointly or separately remove material from a hill top and redeposit it in a valley bottom, truncating the soil profile upslope while burying parts of the lower slope and valley bottom with **colluvium** (moved by gravity), **loess** (moved by wind), or **alluvium** (moved by flowing water). In these lower areas, the new deposits may bury a pre-existing soil, resulting in a **paleosol**, while the newer deposits will themselves eventually experience soil-forming processes, so that there is more than one A horizon in a profile (Fig. 17.13). These processes can be very slow, but

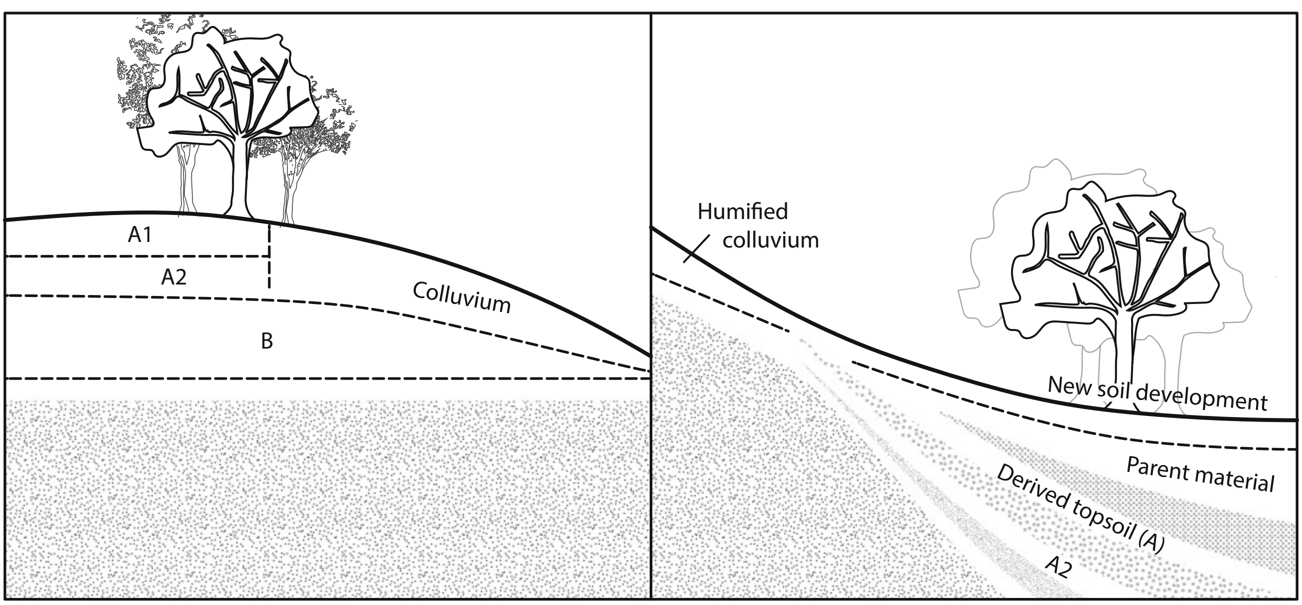


Fig. 17.13 Truncation of a soil profile upslope, burial of downslope soils by colluvium to create paleosols, and new soil development on the colluvium. (Y. Salama, after Butzer 1982: 133)

are sometimes quite fast and dramatic, as in the case of landslides that move huge volumes of sediment, sometimes including artifacts and even small sites, dozens of meters in a few seconds (Field and Banning 1998).

When water transports particles, it carries them along watercourses until the water is too slow to keep them in suspension, so they settle out as alluvium. Typically, alluvium has large, rounded particles where water velocity was high, and fine particles (silts and clays) farther downstream, where water slowed as the channel became broader or less steep (Fig. 17.14). But fast-moving water also cuts into previously deposited alluvium or colluvium and carries away its particles, incising a channel that may have a levee on each bank that results from the settling of particles when water overflows the banks. Typically, the particle sizes are coarser on the levee, and grade into silts and clays with distance away from the stream. This process often results in river channels that are higher than the floodplain on either side (Brown 1997: 17–25).

Size, rounding, sorting, and orientation of particles are good clues to the kind of transport that created sediments (Friedman 1961; Mason and Folk 1958; Sneed and Folk 1958; Udden 1898; Visher 1969). As artifacts are also particles, their orientation, edge damage and size sorting can also provide clues to the processes that removed and transported them (Schiffer 1987: 267–79; Shea 1999).

Archaeologists sometimes describe processes that act on archaeological deposits as "disturbance" because they displace artifacts from what we like to think was their original places of deposition. However, it is not at all clear how we could accurately define the "original" location of most artifacts, while some archaeologists have found it more productive to view archaeological deposits as dynamic, rather than static, entities for which "disturbance" is an inappropriate term. Schiffer (1987) refers to the natural and cultural processes that have acted or continue to act on deposits as "N-transforms" and "C-transforms," respectively. There may also be interplay between these as, for example, human activity, such as forest clearance, can cause water tables to fall or erosion to increase, which can lead to substantial changes to landscapes.

Some of these processes act on small scales. Frost (cryoturbation), growing plants and burrowing animals (**bioturbation**) can all affect the preservation of archaeological

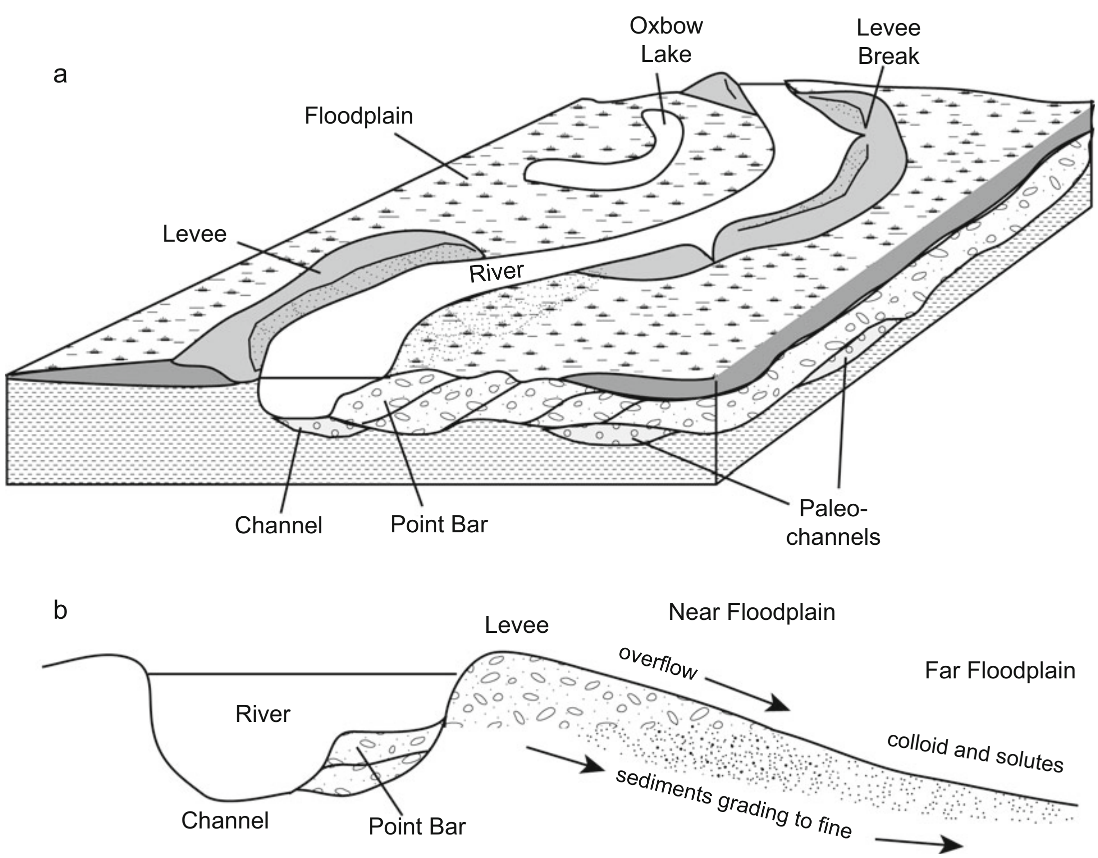


Fig. 17.14 Sediment transport by a river and creation of levees on floodplains (**a**) and cross-section showing particle sorting away from levees (**b**), with vertical exaggeration. (Modified from Courty et al. 1989: 87)

features and the movement of artifacts through capillary action and freezing, displacement by roots, or removal by burrowing, which can change the spatial distributions, sort artifacts by size, and even mimic artificial structures (Limbrey 1975; Rolfsen 1980; Schiffer 1987: 207–215; Wood and Johnson 1978: 344–346). Already, Darwin (1882: 178–231) demonstrated how earthworms could bury ancient Roman pavements in a surprisingly short time, while also undermining pavements and walls to cause subsidence. Ants and termites (Araujo 2013; McBrearty 1990; Robins & Robins 2011), clams (Shafer et al. 1997), crayfish, mole rats (Blackham 2000), rabbits (Pelletier et al. 2017), snails, voles and worms (Canti 2003) can bring older artifacts and bone fragments up to younger surfaces or introduce recent ones into deeper layers, and even concentrate them in ways that we could mistake for cultural activity areas. Tree falls can also bring huge masses of soil or sediment up to the surface along with the roots (Rolfsen 1980: Wood and Johnson 1978: 318–333).

17.10 Environmental Interpretation

Evidence for geomorphological processes provides important clues to past environmental conditions (Brown 1997; Hassan 1978; Goldberg 1992; Courty et al. 1989). In fact, it often has the advantage of highlighting local conditions, while pollen cores and global climate models tend to be more relevant to macroregional climate trends.

For example, sequences of alluvial and colluvial episodes in valley bottoms can provide clues to intensity of rainfall and presence of vegetation on nearby slopes. Because channel downcutting and transport of large stones requires high water velocities, well-sorted alluvium with large particle sizes can result from flash floods in the rainy season of a relatively arid climate. Moister environments in which rainfall is fairly evenly distributed may result in healthy slope vegetation, whose roots help to prevent landslides and flash-flooding, while very arid environments with only sparse vegetation on slopes tend to experience severe erosion from slopewash, and landslides that deposit large volumes of colluvium downslope.

17.11 Cultural Interpretation

Many of the processes that affect sediments both in sites and on landscapes are cultural, and this can aid in our interpretation of archaeological deposits (Butzer 1982: 78; Hassan 1978; Holliday 1992; Rosen 1986; Schiffer 1987: 288–291; Stein 1990). For example, a rather random distribution and orientation of particles in a pit or between walls of a structure might signal an anthropogenic fill deposit, or size sorting at a small scale may even demonstrate that the fill was dumped in one basketful at a time. Orientation and size distributions of stones and artifacts can also help us identify the surfaces that people once walked on; stones and artifacts tend to be "flat-lying" on such surfaces while trampling and sweeping can cause particle sizes on floors to be smaller than in fills (Nielsen 1991; Simms 1988). Sediment characteristics can also help us distinguish mud or mud-brick walls from the debris of decayed mud brick that surrounds them, or to distinguish genuine activity areas from artifact concentrations made by non-human animals.

17.12 Stratigraphic Associations

Sediment analysis can also help us determine whether stratigraphic contexts in different excavation areas of a site are likely to be parts of the same lithostratigraphic layer. If the sediments were deposited at the same time, by the same process, they may be closely similar in their characteristics, such as color, texture, particle roundness and sorting. Such similarity does not guarantee that they are from the same layer, but in combination with other evidence, such as stratigraphy, radiocarbon dates or artifact content, could indicate a strong probability.

Archaeologists have even used such associations at the landscape scale to identify ancient land surfaces. For example, Stafford (1992) reconstructed the topography of a buried Early Woodland surface around the Ambrose Flick site in the upper Mississippi Valley by plotting the elevations of density peaks of micro-artifacts from auger sampling.

Case Study

Microrefuse in Sediments

Fladmark (1982) pioneered the examination of microdebitage, microscopic lithic debris—typically particles in the range of 0.5–5 mm in length—in site sediments as an aid to identifying activity areas, a method that other archaeologists only rarely took up and expanded to other categories of micro-remains in the years that followed (e.g., Hull 1987; Metcalfe and Heath 1990; Rosen 1986; Simms and Heath 1990; Vance 1987). One could make the case, with caution, that these micro-remains are less susceptible than large artifacts to displacement by sweeping and other site-maintenance activities (Simms and Heath 1990: 804; Ullah et al. 2015: 1240), which would seem to make them ideal for obtaining a sense of the spatial structure of routine or habitual activities over an extended time. However, most archaeologists have considered the workload of microrefuse analysis too daunting.

Ullah et al. (2015) demonstrate a sampling strategy that they argue protects the quality of microrefuse data while drastically reducing the time and effort required to carry it out. They gridded the surfaces of Neolithic house floors in the field and collected sediment to a depth of about 1 cm in each grid unit, each yielding a little over a liter of sediment. Back in the lab, they use flotation and nested screens to isolate particles in several increments between 1 and 4 mm, but with particular focus on the 1–1.4 and 1.4–2.0 mm size fractions.

Rather than having a single analyst take on the daunting task of counting micro-artifacts in all of the sample elements or, worse yet, having different analysts count the entirety of different sample elements across house floors (with resultant uncontrolled inter-observer error), they had teams of volunteers do the counting under binocular microscopes (Fig. 17.15). To control for inter-observer effects, they trained and supervised the volunteers, had a reference collection close at hand and, importantly, had every volunteer count a subsample (with replacement) from every grid square on a house floor. Subsampling at one of the sites they sampled this way employed sample elements of 3.0 ml, which was sufficient to obtain reasonable numbers of at least the most common classes of microrefuse, such as flint microflakes and pottery chips. The recording forms are structured in such a way that no volunteer is aware of the counts of other volunteers, as it is important that their observations are independent. After eliminating major outliers (results of volunteers whose counts are consistently outside the 90% confidence interval for a particular class of microrefuse), they use trimmed means or medians to mitigate any remaining interobserver differences. To ensure adequate sample size, the project employs sequential sampling, adding further volunteers until the three-point local slope in the cumulative relative Standard Error is less than 0.03 for three consecutive measures, in other words, until the curve flattens out (Fig. 17.16). Generally, this required sample sizes less than 20 for the most abundant kinds of microrefuse.

Fig. 17.15 Students examining sediment samples for microrefuse

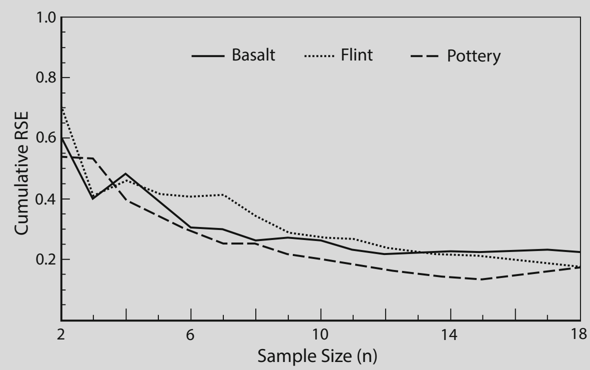


Fig. 17.16 Graph of Relative Standard Error against sample size for sequential sampling. (Y. Salama, after Ullah et al. 2015)

Later spatial analysis of the microrefuse distributions both by their individual densities and by multivariate analysis of the combinations of microrefuse types may allow plausible reconstruction of the spatial organization of activities like grinding, flint-tool maintenance, cooking, and sweeping. In the case of the Late Neolithic site of Tabaqat al-Bûma, Jordan, that included likely cooking activities near a hearth, grinding activities near a large mortar, and flint retouching near a probable doorway or window.

17.13 Summary

- The physical characteristics of sediments are keys to identifying the site-formation processes involved in accumulation and removal of artifacts and other cultural evidence
- Sediment characteristics also help archaeologists identify ancient surfaces and boundaries between depositional events, and to infer which layers in non-contiguous excavation areas may be portions of the same lithostratigraphic unit
- Soil formation, because it takes a long time, helps us identify periods of landscape stability as well as ones of climate change or other impacts on local environments, such as deforestation
- Sedimentary sequences are also good indicators of local environmental change
- Some characteristics of soils and sediments have significant impacts on the probability that certain materials, such as bone, will survive burial
- Others, such as phosphate concentrations, can help us identify settlement or midden areas and testify to the past presence of materials that have since decayed away
- Fragments of artifacts are also clasts, in a sense, and microscopic cultural debris (microrefuse) in some kinds of sediments can be valuable clues to human activities

References Cited

Allen, T. (1968). *Particle size analysis*. London: Halstead Press.

Araujo, A. G. M. (2013). Bioturbation and the upward movement of sediment particles and archaeological materials: Comments on Bueno et al. *Journal of Archaeological Science, 40*, 2124–2127.

ASTM. (2011). *ASTM D2487–17. Standard practice for classification of soils for engineering purposes (Unified soil classification system)*. West Conshohocken: ASTM International.

Blackham, M. (2000). Distinguishing bioturbation and trampling using pottery sherd measures, Tell Fendi, Jordan. *Geoarchaeology, 15*(5), 469–497.

Brown, A. G. (1997). *Alluvial geoarchaeology: Floodplain archaeology and environmental change*. Cambridge: Cambridge University Press.

BSI. (1999). *Code of practice for site investigations, BS 5930*. London: British Standards Institution.

Bullock, P., Federoff, N., Jongerius, A., Stoops, G., & Tursina, T. (1985). *Handbook for soil thin section description*. Wolverhampton: Waine Research Publications.

Butzer, K. W. (1982). *Archaeology as human ecology*. Cambridge: Cambridge University Press.

Butzer, K. W. (2008). Challenges for a cross-disciplinary geoarchaeology: The intersection between environmental history and geomorphology. *Geomorphology, 101*(1–2), 402–411.

Canti, M. G. (2003). Earthworm activity and archaeological stratigraphy: A review of products and processes. *Journal of Archaeological Science, 30*(2), 135–148.

Courty, M. A., Goldberg, P., & Macphail, R. (1989). *Soils and micromorphology in archaeology*. Cambridge: Cambridge University Press.

Cruz-Matías, I., Ayala, D., Hiller, D., Gutsch, S., Zacharias, M., Estradé, S., & Peiro, F. (2019). Sphericity and roundness computation for particles using the extreme vertices model. *Journal of Computational Science, 30*, 28–40.

Darwin, C. (1882). *The formation of vegetable mould, through the action of worms with observations on their habits*. London: John Murray.

Davidson, D. A. (1973). Particle size and phosphate analysis—Evidence for the formation of a tell. *Archaeometry, 15*, 143–152.

Durian, D. J., Bideaud, H., Duringer, P., Schroder, A., Thalmann, F., & Marques, C. M. (2006). *What is in a pebble shape?* http://arxiv.org/abs/cond-mat/0607061v1.

Field, J., & Banning, E. B. (1998). Hillslope processes and archaeology in Wadi Ziqlab, Jordan. *Geoarchaeology, 13*, 595–616.

Fladmark, K. R. (1982). Microdebitage analysis: Initial considerations. *Journal of Archaeological Science, 9*, 205–220.

Folk, R. L. (1966). A review of grain size parameters. *Sedimentology, 6*, 73–93.

Fouache, E. (2013). The geoarchaeological approach. In C. Corsi et al. (Eds.), *Good practice in archaeological diagnostics* (pp. 245–252). Cham: Springer.

French, C. (2003). *Geoarchaeology in action: Studies in soil micromorphology and landscape evolution*. London: Routledge.

Friedman, G. M. (1961). Distinction between dune, beach and river sands from their textural characteristics. *Journal of Sedimentary Petrology, 27*, 3–27.

Goldberg, P. (1992). Micromorphology, soils, and archaeological sites. In V. T. Holliday (Ed.), *Soils in archaeology* (pp. 145–167). Washington: Smithsonian Institution Press.

Goldberg, P., & Macphail, R. I. (2006). *Practical and theoretical geoarchaeology*. Oxford: Blackwell Science.

Goldberg, P., Nash, D. T., & Petraglia, M. D. (Eds.). (1993). *Formation processes in archaeological context* (Monographs in world prehistory) (Vol. 17). Madison: Prehistory Press.

Hassan, F. A. (1978). Sediments in archaeology: Methods and implications for paleoenvironmental and cultural analysis. *Journal of Field Archaeology, 5*, 197–213.

Holliday, V. T. (Ed.). (1992). *Soils in archaeology. Landscape evolution and human occupation*. Washington: Smithsonian Institution Press.

Holliday, V. T. (2004). *Soils in archaeological research*. Oxford: Oxford University Press.

Hull, K. L. (1987). Identification of cultural site formation processes through microdebitage analysis. *American Antiquity, 52*, 772–783.

Inman, D. L. (1962). Measures for describing the size distributions of sediments. *Journal of Sedimentary Petrology, 22*, 125–145.

Krumbein, W. C. (1934). Size frequency distribution of sediments. *Journal of Sedimentary Petrology, 4*, 65–77.

Krumbein, W. C. (1941). Measurement and geological significance of shape and roundness of sedimentary particles. *Journal of Sedimentary Research, 11*, 64–72.

Limbrey, S. (1975). *Soil science and archaeology*. New York: Academic.

Mason, C., & Folk, R. L. (1958). Differentiation of beach, duen and Aeolian flat environments by size analysis. *Journal of Sedimentary Petrology, 28*, 212–226.

McBrearty, S. (1990). Consider the humble termite: Termites as agents of post-depositional disturbance at African archaeological sites. *Journal of Archaeological Science, 17*(2), 111–143.

Metcalfe, D., & Heath, K. M. (1990). Microrefuse and site structure: The hearths and floors of the heartbreak hotel. *American Antiquity, 55*, 781–796.

Munsell Color. (2010). *Munsell soil color charts: With genuine Munsell color chips*. Grand Rapids: Munsell Color.

NACSN. (1983). North American commission on stratigraphic nomenclature: North American Stratigraphic Code. *American Association of Petroleum Geologists Bulletin, 67*, 841–875.

Nielsen, A. E. (1991). Trampling the archeological record: An experimental study. *American Antiquity, 56*, 483–503.

Pelletier, M., Royer, A., Holliday, T. W., Discamps, E., Madelaine, S., & Maureille, B. (2017). Rabbits in the grave! Consequences of bioturbation on the Neanderthal 'burial' at Regourdout (Montignac-sur-Vézère, Dordogne). *Journal of Human Evolution, 110*, 1–17.

Rapp, G. R., & Hill, C. L. (2006). *Geoarchaeology: The earth science approach to archaeological interpretation*. New Haven: Yale University Press.

Rice, E. W., Baird, R. B., & Eaton, A. D. (Eds.). (2017). *Standard methods for the examination of water and wastewater* (23rd ed.). American Public Health Association.

Rittenhouse, G. (1943). A visual method of estimating two-dimensional sphericity. *Journal of Sedimentary Pedology, 13*, 79–81.

Robins, R., & Robins, A. (2011). The antics of ants: Ants as agents of bioturbation in a midden deposit in south-East Queensland. *Environmental Archaeology, 16*(2), 151–161.

Rolfsen, P. (1980). Disturbance of archaeological layers by processes in the soil. *Norwegian Archaeological Review, 13*, 110–118.

Rosen, A. M. (1986). *Cities of clay: The geoarchaeology of tells*. Chicago: University of Chicago Press.

Schiffer, M. B. (1987). *Formation processes of the archaeological record*. Albuquerque: University of New Mexico Press.

Schwartz, G. T. (1967). Prospecting without a computer in southern Switzerland. *Estratto da prospezioni Archaeologische, 2*, 73–80.

Shackley, M. L. (1975). *Archaeological sediments: A survey of analytical methods*. London: Butterworth.

Shafer, B. S., Dering, J. P., Labadie, J., Pearl, F. B., & Huebner, A. M. (1997). Bioturbation of submerged sites by the Asiatic clam: A case study from Amistad Reservoir, SW Texas. *Journal of Field Archaeology, 24*(1), 135–138.

Shea, J. J. (1999). Artifact abrasion, fluvial processes and "living floors" from the early Paleolithic site of 'Ubeidiya (Jordan Valley, Israel). *Geoarchaeology, 14*, 191–207.

Simms, S. R. (1988). The archaeological structure of a Bedouin camp. *Journal of Archaeological Science, 15*, 197–211.

Simms, S. R., & Heath, K. M. (1990). Site structure of the orbit inn: An application of ethnoarchaeology. *American Antiquity, 55*, 797–813.

Sneed, E. D., & Folk, R. L. (1958). Pebbles in the lower Colorado River, Texas: A study in particle morphogenesis. *Journal of Geology, 66*, 114–150.

Soil Survey Staff. (1975). *Soil taxonomy: A basic system of soil classification for making and interpreting soil surveys*. Washington: Department of Agriculture.

Stafford, C. R. (1992). Archaeological stratigraphy and paleotopography. In C. R. Stafford (Ed.), *Early woodland occupations at the Ambrose flick site in the northern Sny bottom of west-Central Illinois* (pp. 58–94). Kampsville: Center for American Archeology.

Stein, J. K. (1987). Deposits for archaeologists. *Advances in Archaeological Method and Theory, 2*, 319–396.

Stein, J. K. (1990). *Deciphering a shell midden*. San Diego: Academic.

Stein, J. K., & Farrand, W. R. (2001). *Sediments in archaeological context*. Salt Lake City: University of Utah Press.

Thien, S. J. (1979). A flow diagram for teaching texure by feel analysis. *Journal of Agronomic Education, 8*, 54–55.

Udden, J. A. (1898). *The mechanical composition of wind deposits* (Augustana library publication no. 1). Rock Island: Lutheran Augustana Book Concern.

Ullah, I., Duffy, P., & Banning, E. B. (2015). Modernizing spatial micro-refuse analysis: New methods for collecting, analyzing, and interpreting the spatial patterning of micro-refuse from house-floor contexts. *Journal of Archaeological Method and Theory, 22*(4), 1238–1262.

Vance, E. D. (1987). Microdebitage analysis in activity analysis: An application. *Northwest Anthropological Research Notes, 20*, 179–189.

Visher, G. S. (1969). Grain size distributions and depositional processes. *Journal of Sedimentary Research, 39*, 1074–1106.

Wadell, H. (1932). Volume, shape, and roundness of rock particles. *The Journal of Geology, 40*, 443–451.

Wadell, H. (1933). Sphericity and roundness of rock particles. *The Journal of Geology, 41*, 310–331.

Waters, M. R. (1992). *Principles of geoarchaeology: A north American perspective*. Tucson: University of Arizona Press.

Wentworth, C. K. (1922a). The shape of pebbles. *Bulletin of the US Geological Survey, 730-C*, 91–114.

Wentworth, C. K. (1922b). A scale of grade and class terms for clastic sediments. *The Journal of Geology, 30*, 377–392.

Wood, W. R., & Johnson, D. L. (1978). A survey of disturbance processes in archaeological site formation. *Advances in Archaeological Method and Theory, 1*, 315–381.

18 Seriation

In dating any class of objects, such as spears, adzes, fibulae, combs, etc., the process is to look out the age or the range of age of each of the graves in which such objects are found.

Petrie (1899: 299)

This chapter is the first of several that deal explicitly with the chronological dimension of archaeological analysis. Seriation involves methods by which archaeologists attempt to put assemblages of artifacts into a serial order, that is, to create an ordinal chronological scale. Seriation orders units—assemblages, artifacts, or other entities—along a single dimension, assumed to be time, so that adjacent units are more similar to one another than to nonadjacent units. Some of the grouping methods already discussed in Chap. 3 can help accomplish this.

These and the methods described below depend on the characteristics of the artifacts found in assemblages and on the assumption that changes in technology, style, and sources of exotic artifacts cause the character of artifact populations to differ over time (Baxter 2003; Laxton 1990; Lyman and O'Brien 2006; Marquardt 1978; McCafferty 2008; O'Brien and Lyman 2002).

Although numismatists had already been using a version of seriation to analyze coin hoards since Bartolomeo Borghesi introduced it about 1820 (Crawford 1990), the classic early archaeological use of seriation was Flinders Petrie's (1899) "sequencing" of predynastic Egyptian graves (Fig. 18.1). Petrie informally used what archaeologists call the **concentration principle** to arrange 900 graves into a serial order on the basis of 804 types and varieties of pottery and other artifacts. The concentration principle is a preference for arrangements that minimize the ranges of varieties over the sequence, in other words, that make the types "clump" in the sequence. Other superficially similar methods, such as those of General Pitt-Rivers (1875), ordered artifacts instead by their assumed evolutionary development.

18.1 Incidence and Frequency Seriation

The simplest seriation, **incidence seriation**, uses a dichotomous model inspired by Petrie's "sequence dating" but formalized much later. This involves recording the presence (1) or absence (0) of each type instead of its relative abundance and following the rule that, once a 0 follows a 1, it can only be followed by more zero values. For example, you could have 0, 0, 1, 1, 0, 0, but not 0, 1, 0, 1, 0, 0. This is the concentration principle: the ones must cluster together with no intervening zeroes (Table 18.1). In honor of Petrie, a series ordered in this way is called a **P-matrix**. In some instances, two or three contexts are interchangeable—their order does not affect the concentration principle—so that we can only treat them as contemporaneous.

It is more common for archaeologists to use **frequency seriation**, based on a model that describes the way the relative abundances of "types," or classes of artifact, not just their presence or absence, are expected to change over time. For typologically-based seriation, the key assumption of this model—the **Kendall model**—is that each artifact type, once introduced, grows in its relative abundance (incipience), eventually reaches a peak (fluorescence), and then declines as other new, competing artifact types begin to displace it (Fig. 18.2; Buck et al. 1996: 328; Dempsey and Baumhoff 1963; Kendall 1971; Phillips et al. 1951; Robinson 1951). If drawn as a bar graph with the bars arranged horizontally and centered, this pattern forms a "battleship curve" (Fig. 18.3; Ford 1962). Because the data are proportions or percentages, the various types are not independent of one another (when one goes up, something else has to go down in relative frequency).

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47992-3_18) contains supplementary material, which is available to authorized users.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_18

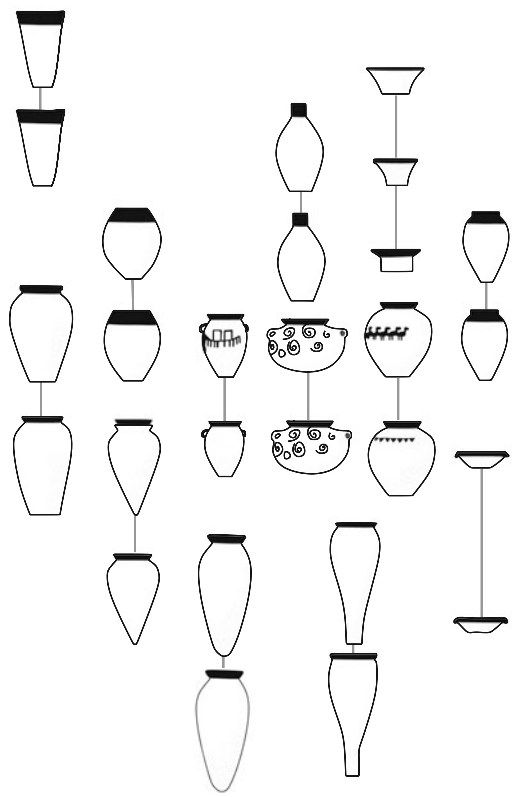


Fig. 18.1 Abbreviated version of one of Petrie's seriations, which Petrie labelled numerically with so-called "sequence numbers" at right, and line segments showing continuity of types. (Y. Salama, modified from Petrie 1899: fig. 1)

Table 18.1 An incidence matrix (left) showing the presence (1) or absence (0) of each of four artifact types A-D among eight different archaeological contexts; and (right) a seriation of those contexts by the concentration principle to create a "P-matrix"

	Artifact type					Artifact type			
Context	A	B	C	D	Context	A	B	C	D
1	1	0	1	0	2	0	0	0	1
2	0	0	0	1	8	0	1	0	1
3	1	1	0	1	6	1	1	0	1
4	1	1	0	0	3	1	1	0	1
5	0	0	1	0	4	1	1	0	0
6	1	1	0	1	7	1	0	1	0
7	1	0	1	0	1	1	0	1	0
8	0	1	0	1	5	0	0	1	0

From Kendall (1971: 220)

Notice how the "1" values are now clustered, with no intervening zeroes

Other assumptions of this model are that each unit corresponds with a brief and comparable period of time, that the content of each unit is a representative sample of a population of artifacts that were in use during that time and place (see Chap. 6), that all units belong to the same cultural

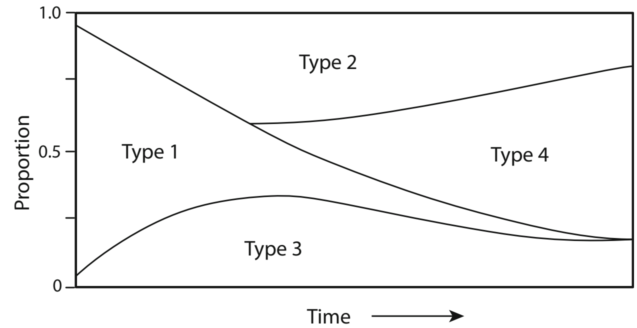


Fig. 18.2 Kendall's model for the way the proportions of four types of artifact would vary over time. Each type has a unimodal distribution such that it gradually increases, reaches a peak, and then decreases in abundance. (After Doran and Hodson 1975: 272)

tradition, and that all the units come from a reasonably small region (Dunnell 1970: Marquardt 1978: 261–297; O'Brien and Lyman 2002: 117–119). The last two assumptions are to ensure that the dimension along which we order the units is time, rather than space or cultural affiliation, while the assumption of a representative sample is an important one whose implications merit further attention, below. Further implicit assumptions, not always recognized, are that the populations from which the artifacts came are themselves serially ordered, and the units are samples from different populations. In other words, none of the assemblages are exactly contemporary. As you can imagine, many of these assumptions pose difficulties, as our imposition of an "episodic" model—treating each assemblage as a sort of glimpse of time—ignores the likelihood that assemblages accumulated over somewhat extended and somewhat different intervals of time.

Mathematically, we can represent the relative abundances of types in a matrix of m rows and n columns, where m is the number of archaeological units (layers, graves, components, etc.) and n is the number of artifact types. Each cell in the matrix contains a value, $a_{i,j}$, meaning the jth entry in the ith row, or the number of artifacts of type j in the ith archaeological unit. We conclude that the matrix is seriated if the numbers going down each column, once they have decreased, never rise again. We refer to this property of never increasing again as decreasing **monotonically**. For example, in the left side of Table 18.2, the archaeological contexts are in arbitrary order but, on the right, reordering of the same contexts results in the relative abundance of each type never increasing once it has started to decrease, a **Q-matrix**.

Although the seriation arranges the contexts in a linear order (4, 1, 7, 5, 3, 8, 2, 6), the direction of the ordering is arbitrary. Except by reference to some other data, we cannot rule out that the order 6, 2, 8, 3, 5, 7, 1, 4 is the correct one.

There are, however, problems with the Kendall method (McNutt 2005). First, it is not always true that in correctly ordered populations of artifacts the relative abundance of a

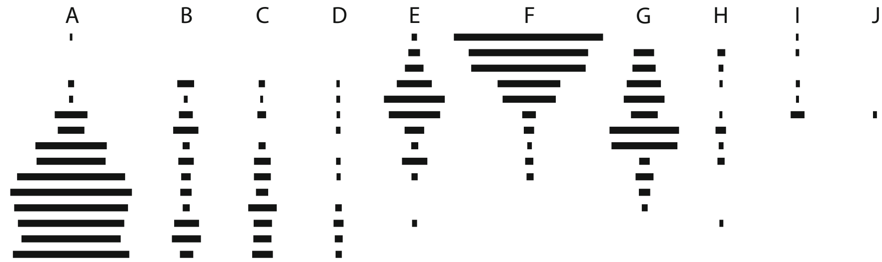


Fig. 18.3 Example of a "battleship curve" for Allegheny Region pottery types labelled A through J (Y. Salama, after Ford 1962: fig. 9). Each horizontal row of bars represents pottery frequencies from a single assemblage. Note that it is not necessary to center the bars in order to seriate the assemblages correctly, as the key feature is that each type increases, reaches a peak, then declines in relative frequency

Table 18.2 On the left, a matrix with the percentages of four different artifact types in eight different contexts, in arbitrary order. On the right, the same eight contexts seriated so that the percentages never increase again after they begin to decrease (a Q-matrix, after Kendall 1971: 219)

	Artifact type					Artifact type			
Context	A	B	C	D	Context	A	B	C	D
1	10	20	1	70	4	0	10	0	90
2	20	0	70	10	1	10	20	1	70
3	40	10	30	20	7	30	30	5	35
4	0	10	0	90	5	40	20	15	25
5	40	20	15	25	3	40	10	30	20
6	10	0	90	0	8	30	5	50	15
7	30	30	5	35	2	20	0	70	10
8	30	5	50	15	6	10	0	90	0

type will never increase after it has decreased. More realistically, there are fluctuations that the Kendall model does not accommodate. Even Ford's (1962) examples of battleship curves, such as those for surface collections in Virginia (Fig. 18.3), show numerous fluctuations that deviate from the Kendall model, and such anomalies persist in more recent applications (e.g., Wesler 1999). Second, since any particular context is only a sample of a population, and probably not a random one at that (Madsen 1988), typical archaeological cases may violate the assumption of representative sampling or just suffer from insufficient sample size; it is easy to imagine cases where a particular sample could omit an artifact type that was, in reality, fairly common at the time of assemblage formation. The Kendall model makes no allowance for sampling error, preservation issues and other factors that create noise in the data (Buck and Sahu 2000). Third, the algorithms for finding the best ordering of units do not account for the possibility that there is more than one possible order, including some that are almost as likely as the "best" one (Buck and Litton 1991; Buck et al. 1996: 329). That is why most researchers have favored a probabilistic approach to seriation—finding the best, yet imperfect, fit to the Kendall model, rather than a deterministic one that assumes there is only a single solution. One reason that such alternative orders are possible is that some of the samples could be from the same population, or from two contemporary populations.

One way to address the first of these problems is to turn to the overall similarity between units, rather than assuming that each and every type will increase to a peak and then decrease monotonically. To do this, we can make a new matrix that is n x n instead of n x m, that is, with archaeological contexts along both axes (Table 18.3). For each pair of contexts, we then measure the degree of similarity by Robinson's (1951) Index of Agreement (IA):

$$IA_{jk} = 200 - \left(\sum_{i=1}^{m} \left| x_{ji} - x_{ki} \right| \right)$$

or the sum of the absolute values of the difference between the percentages of each type for each pair of archaeological units (j and k), subtracted from 200. In other words, if two archaeological contexts have exactly the same distribution of types, the summed differences between them (a measure of dissimilarity) will be 0 and the IA will be 200. If, on the other hand, the two contexts are maximally different (i.e., when one has 100% of one type and the other has 100% of another type), the summed differences will be 200% and IA will be 0. Subtracting from 200 simply turns a dissimilarity coefficient into a similarity coefficient.

Returning to the data from the matrix in Table 18.2, the absolute values of differences between the four types for contexts 1 and 2 are 20-10, 20-0, 70-0, and 70-10. The sum of these differences is 10 + 20 + 70 + 60 = 160, so that IA = 200 − 160 = 40. For contexts 1 and 3, they are 30 + 10 + 30 + 50 = 120, so IA = 80. We can continue this process for each pair of contexts to produce the matrix in Table 18.3. We only have to fill out half the matrix, as it is perfectly symmetrical about the diagonal, where all values of IA are 200.

The next step is to rearrange the rows and columns so that higher IA values cluster near the diagonal and IA decreases toward the upper right corner. Some of the highest values are 170 for the pair (3,5) and 160 for (5,7). Meanwhile, the pair (4,6) shows no agreement at all (IA = 0), indicating that they should be at opposite ends of the sequence. This kind of reasoning provides the basis for ordering the contexts as in Table 18.4.

Dempsey and Baumhoff's (1963: 499) variation on Robinson's IA reduces the data to a dichotomous scale "so

Table 18.3 Matrix of Robinson's Indices of Agreement (IA) for the data in Table 18.2

Context	1	2	3	4	5	6	7	8
1	200	40	80	140	110	20	130	60
2		200	120	20	110	160	70	160
3			200	60	170	80	130	160
4				200	50	0	90	40
5					200	50	160	130
6						200	30	125
7							200	110
8								200

Table 18.4 Rearranged matrix of the Robinson's Indices of Agreement (IA) for the data from Tables 18.2 and 18.3. In this case, the same order results as in the right half of Table 18.2. For the most part, the IA values fall off with distance from the diagonal

Context	4	1	7	5	3	8	2	6
4	200	140	90	50	60	40	20	0
1		200	130	110	80	60	40	20
7			200	160	130	110	70	30
5				200	170	130	110	50
3					200	160	120	80
8						200	160	125
2							200	160
6								200

that the presence (or absence) of any one type contains no necessary implication concerning the presence (or absence) of any other type." In other words, it counters the problem of inter-type dependence in the Kendall model. Their approach uses similarity coefficients in the same way as in clustering (pp. 30–32). Although it solves the problem of interdependence of proportions, it has the usual problem of presence/absence data: a single artifact of one type has just as much weight as hundreds of examples of another type (Marquardt 1978: 268). Dempsey and Baumhoff (1963: 498) argue that this is an advantage as rare types may be chronologically "diagnostic."

Marquardt (1978) reviews many other variations on and alternatives to the Robinson (1951) method. Some interesting ones are Cowgill's (1972), which measures similarity, not between pairs of contexts, but pairs of types, Wilkinson's (1974), which treats artifacts and contexts equally, and LeBlanc's (1975), which considers distributions of attributes, rather than types.

18.2 Multidimensional Scaling (MDS) and Correspondence Analysis (CA)

Just as we could use similarity matrices to group artifacts into types, or to seriate contexts or artifacts as described in the last section, we can also use some of the grouping methods that are based on these matrices. Although the methods are identical, there is a philosophical difference in the application of these methods to either contexts or artifacts. In the former case, we are ordering assemblages, much as with the previous methods and employing many of the same assumptions. In the latter, there is an implicit further assumption of "descent with modification" (Darwin 1859; cf. O'Brien and Lyman 2002), much as with Pitt-Rivers' artifact sequences.

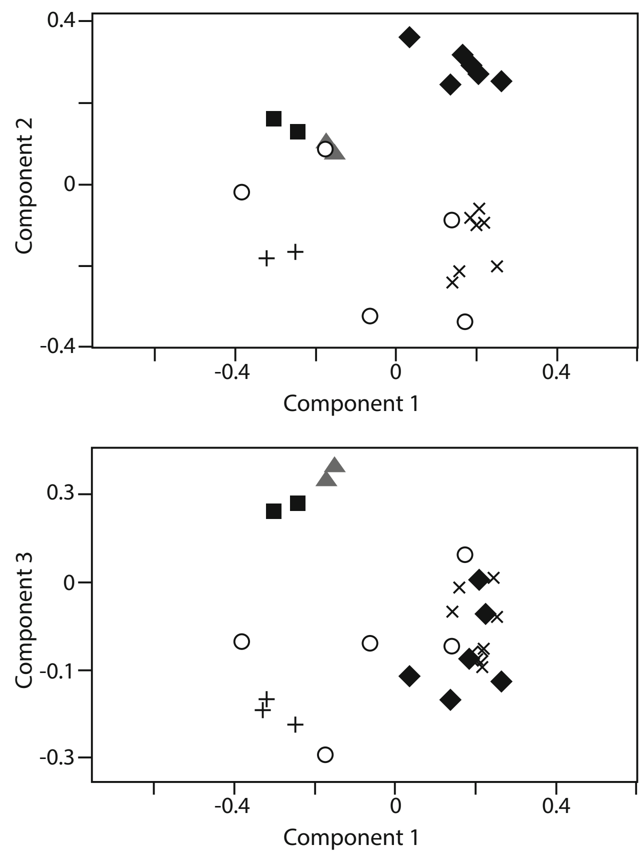


Fig. 18.4 Component plots for a non-metric MDS of a matrix of 19 petrographic characteristics in the fabric of 25 pottery samples from Can Sora, Spain. (Modified from Baxter 2003: 88)

Multidimensional scaling (MDS, p. 33) creates something like a "map" that represents dissimilarities among units (contexts or artifacts) as distances in space. Because the resulting space is multidimensional (one dimension for each attribute or type), representing this "map" on a two-dimensional image requires distortion of distances, and MDS permits these distortions as long as they preserve the rank-order of the distances. Often, the resulting "map" has a near-circular or horseshoe shape (Fig. 18.4) that results from having many units with almost the same dissimilarity with some other unit. If time is the principal dimension along which the units are distributed, then the oldest context or artifact is at one end of the horseshoe and the youngest at the other, so that the units are chronologically ordered.

Correspondence analysis (CA) is an alternative that, somewhat like multidimensional scaling, often yields a

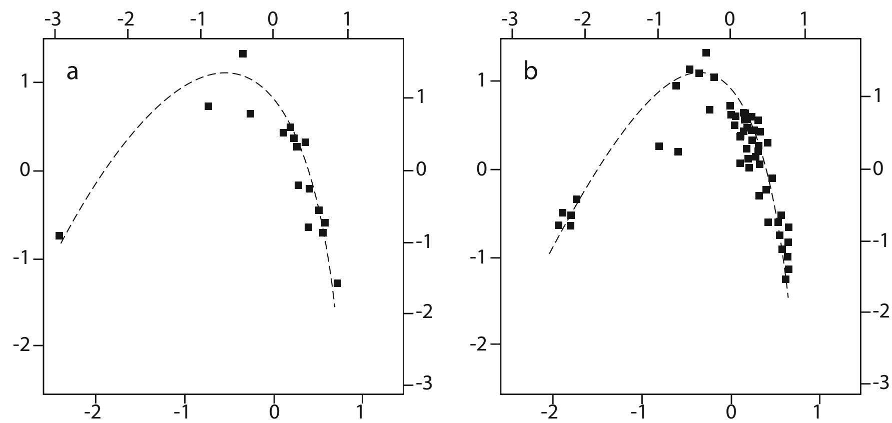


Fig. 18.5 Correspondence Analysis of 52 pottery types in 16 tomb assemblages at the site of Tell al-Farah al-Janubiyyah (after Baxter 2003: 137), with results by tombs (**a**) and pottery types (**b**). Notice how points tend to form horseshoe-shaped curves (dashed), indicating an approximate (chronological) order

U-shaped string of data points in a scatterplot (Fig. 18.5). Unlike the Kendall and Robinson methods, CA does not require the assumption that types first increase and then decrease monotonically in abundance. Consequently, it can, and often does, produce a quite different ordering of the data. However, CA does assume independence between the artifacts and the contexts, something the Kendall-Robinson models do not require, and the presence of many zero values in typical archaeological data tables is a problem for this assumption (Buck and Sahu 2000).

Again, if time is the main contributor to the ordering, and the types or attributes being counted are chronologically sensitive, then the oldest units will be at one end of the curve, and the youngest at the other. Where the distribution of points is not U-shaped, one might take the first axis of the CA as the best seriation order, but most archaeologists consider the "horseshoe" to be a good indicator of a successful seriation. Today, CA is one of the most common ways to "seriate" artifacts and contexts, if only in an exploratory fashion (Baxter 2003: 204–206; Shennan 1997: 342).

18.3 Bayesian Approaches to Seriation

The similarity-matrix and multivariate methods just described provide a single "best" unidimensional order for the units, but real data can be "noisy" and the "best" order ignores the possibility that there could be other orders that are almost as likely (Baxter 2003: 207; Buck et al. 1996: 329; Buck and Sahu 2000; Halekoh and Vach 2004).

Other Bayesian research in this vein has included Buck and Sahu's (2000) introduction of a hierarchical Bayesian model for CA, and Halekoh and Vach's (2004) attempt to seriate graves from the important La Tène cemetery of Münsingen-Rain, near Bern, Switzerland, by using a stochastic model of the unimodal distribution of types (the concentration principle) in an incidence matrix.

Despite their potential, Bayesian approaches to seriation have yet to gain wide acceptance, and Baxter (2003: 207) notes that such acceptance will require more experience with large, realistic data sets.

18.4 A New Approach to Deterministic Seriation

Lipo et al. (2015) have noted that probabilistic methods for seriation, such as MDS or CA, simplify data sets in a way that omits a lot of information and makes it difficult to determine the validity of the resulting order. Meanwhile, deterministic methods become unmanageable, even with modern computers, when there are more than about 14 assemblages to be ordered because there are too many possible permutations of assemblage orders to evaluate in a reasonable time. Their solution is the Iterative Determinitive Seriation Solutions (IDSS), which drastically reduces the number of permutations that need to be examined. IDSS entails an iterative procedure, starting with valid sequences of just a few assemblages (valid in the sense of not violating seriation assumptions), then using these as "building-blocks" for longer sequences. This avoids the step of having to calculate all possible combinations of assemblages.

18.5 Die-Linkage of Coins

One very specialized class of artifact—coinage—offers another kind of seriation that would not work for others. Die-linking is based on the fact that most ancient and mediaeval coins were made by striking a flat disk of metal between two dies that wear out after a period of use (see p. 222). Because the upper (reverse) die, which is struck with a hammer, wore or cracked more quickly than the lower (obverse) die, which was set into an anvil, coiners changed the upper and lower dies at different times. In addition, it is likely that coiners kept reverse dies in a secure storage box overnight, and next morning they could be paired with different obverse dies. These factors resulted in the association

Example

Buck and Litton (1991) use an iterative algorithm called a Gibbs sampler, some arbitrary starting values, likelihoods modelled by the multinomial and prior probability by Dirichlet distributions to create seriations that take into account the stochastic nature of real data. For example, site deposits can contain “residual” artifacts (see p. 323) that belong to earlier periods, or some artifact types can be missing from a tomb because of small-sample effects. By carrying out many iterations of this process, each of which may result in a somewhat different ordering, they can discover which order is most probable, and which others fairly probable. Because this takes seriously the possibility that the “best” order might be incorrect, it is an approach that is quite relevant to the theme of quality in data analysis.

After 1000 iterations, using the same data that Laxton and Restorick (1989) used to compare the Kendall method with CA (Table 18.5), they found that the order resulting from the Kendall method (2, 5, 3, 6, 1, 4) has the highest probability but other orders had significant probabilities (Table 18.6). Surprisingly, the other orders found do not include the one that Laxton & Restorick found by CA (3, 6, 5, 2, 1, 4), which would suggest that it is improbable. It is also notable that the three most probable solutions, with a combined probability of 0.893, are very similar, only varying in the order of 1 and 4 or 3 and 6.

In this example, there was no outside information to guide the selection of prior probabilities, but one potentially great advantage of this method is that it can, like any Bayesian method, incorporate information that has a bearing on the correct order. For example, there could be stratigraphic information or radiocarbon dates for some, but not all, of the contexts being ordered, and this would have definite bearing on the “best” outcome by allowing informative priors.

Table 18.5 Fictitious artifact counts (from Laxton and Restorick 1989) that Buck and Litton (1991) use for demonstrating Bayesian seriation

	Artifact type						
Site	1	2	3	4	5	6	7
1	20	3	4	42	18	0	13
2	85	3	12	0	0	0	0
3	26	40	8	0	0	26	0
4	20	1	4	13	58	0	4
5	67	10	23	0	0	0	0
6	26	29	8	3	0	33	1

Table 18.6 Results from 1000 iterations of ordering the six sites in the demonstration of Bayesian seriation by Buck and Litton (1991: 98)

Order	Frequency	Posterior probability
2, 5, 3, 6, 1, 4	679	0.679
2, 5, 3, 6, 4, 1	128	0.128
2, 5, 6, 3, 1, 4	86	0.086
4, 1, 6, 3, 2, 5	46	0.046
2, 5, 6, 3, 4, 1	28	0.028
3, 6, 5, 2, 1, 4	16	0.016
4, 1, 3, 6 2, 5	8	0.008
3, 6 5, 2, 4, 1	4	0.004
5, 2, 3, 6 4, 1	3	0.003
4, 1, 2, 5, 3, 6	1	0.001
5, 2, 6, 3, 4, 1	1	0.001

of each obverse die with two or more reverse dies, while some reverse dies were used with more than one obverse die. And because dies prior to the modern period were engraved by hand, no two dies were exactly alike, even when the engraver intended to reproduce a design very faithfully. Consequently, careful examination of coins to associate them with the various dies, makes it possible to work out a sequence of die use (Fig. 18.6). As with other kinds of seriation, it may not be obvious which end of the sequence is early and which is late, but evidence from increasing die wear or formation of a die crack, either of which would only occur after the die had been used for some time, can help us determine the correct direction of the sequence when there are no other indications, such as a Roman emperor’s titles, to help us do so (Laing 1969: 26–28).

18.6 Quality in the Use and Interpretation of Seriations

Effective use of seriation requires careful attention to assumptions, most notably to have representative samples of artifacts from meaningful contexts that are free of disturbances that could have introduced intrusive or residual artifacts (Marquardt 1978: 292–304; Madsen 1988; McNutt 2005). Any departure from the basic assumptions could result in a seriation along some dimension that is not time, such as space, function, or site-formation process. One way to check to see if the resulting seriation is ordered chronologically is to have at least some chronometric dates, such as radiocarbon assays.

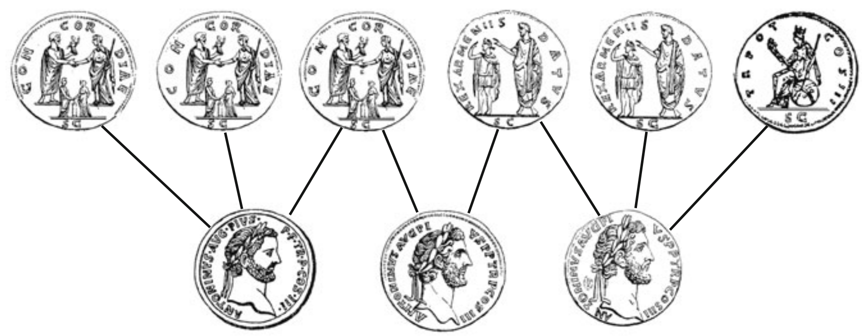


Fig. 18.6 Simplified demonstration of die-linking among coins to establish their relative chronology, using coins of Antoninus Pius (Roman emperor from AD 138 to 161). Because each obverse die was used in combination with several reverse dies, which wore out faster, some obverses share a reverse die, and this provides a basis for ordering the coins from which they were struck. The obverse inscriptions all indicate a date during the emperor's third consulship (AD 140–141)

Scott (1993) additionally notes that archaeologists need to give adequate attention to whether similarity coefficients are based on chronologically sensitive attributes. He worries that giving all contexts equal weight, regardless of the size or nature of their assemblages, could distort results. Scott also notes that, unlike Petrie's old sequence dating, prevalent methods do not help fit new assemblages into previously established sequences. He offers a parametric method that treats the age of each context as a parameter to be estimated. This has the advantages of providing estimates with standard errors, allowing us to judge whether pairs of assemblages are statistically different from one another, and incorporating known dates into the analysis. He models the unimodal variation in the abundance of types over time with a normal distribution, and considers the distribution of artifacts in an assemblage as fitting a Poisson model.

18.7 Summary

- Seriation provides tools to arrange artifacts or archaeological contexts in time on an ordinal scale when evidence from stratigraphy or independent dating is unavailable
- Its key principle, the concentration principle, is that artifact types cluster in time
- Frequency seriation, or the Kendall method, is based on matrices of artifact abundances in different contexts, which may result in an optimal order (Q-matrix with "battleship curves") if all the assumptions are valid
- Incidence seriation, based only on presence or absence of artifact types in each context, results in a P-matrix (after its quasi-originator, Flinders Petrie)
- It is also possible to seriate artifacts whose attributes have "evolved" over time. Although this has often involved an implicit progressivist assumption that was popular in the late nineteenth century but less acceptable today, there are also neo-evolutionary versions of artifact descent through modification
- Multivariate methods such as CA and PCA can order either contexts or artifacts, and it is possible that time is the most influential dimension on that order if most of seriation's assumptions are met
- There are other ordering methods, most notably coinage die-linking, that also seriate specialized classes of artifacts but have principles that do not depend on relative abundance of types or artifacts or the concentration principle

References Cited

Baxter, M. (2003). *Statistics in archaeology*. London: Hodder Arnold.

Buck, C. E., Cavanagh, W., & Litton, C. (1996). *Bayesian approach to interpreting archaeological data*. New York: Wiley.

Buck, C. E., & Litton, C. (1991). A computational Bayes approach to some common archaeological problems. In K. Lockyear & S. Rahtz (Eds.), *CAA91. Computer applications and quantitative methods in archaeology 1990* (pp. 93–99). Oxford: Tempus Reparatum.

Buck, C. E., & Sahu, S. K. (2000). Bayesian models for relative archaeological chronology building. *Journal of the Royal Statistical Society C (Applied Statistics), 49*(4), 423–440.

Cowgill, G. (1972). Models, methods, and techniques for seriation. In D. L. Clarke (Ed.), *Models in archaeology* (pp. 381–424). London: Methuen.

Crawford, M. H. (1990). From Borghesi to Mommsen: The creation of an exact science. In M. H. Crawford, C. R. Ligota, & J. B. Trapp (Eds.), *Medals and coins from Budé to Mommsen* (pp. 125–132). London: Warburg Institute Surveys and Texts 21.

Dempsey, P., & Baumhoff, M. (1963). The statistical use of artifact distributions to establish chronological sequence. *American Antiquity, 28*, 496–509.

Doran, J. E., & Hodson, F. R. (1975). *Mathematics and computers in archaeology*. Edinburgh: Edinburgh University Press.

Dunnell, R. C. (1970). Seriation method and its evaluation. *American Antiquity, 35*, 305–319.

Ford, J. A. (1962). *A quantitative method for deriving cultural chronology* (Technical manual 1, Pan American Union). Washington: Organization of American States.

Halekoh, U., & Vach, W. (2004). A Bayesian approach to seriation problems in archaeology. *Computational Statistics and Data Analysis, 45*(3), 651–673.

Kendall, D. G. (1971). Seriation from abundance matrices. In F. R. Hodson, D. Kendall, & P. Tauta (Eds.), *Mathematics in the* *archaeological and historical sciences* (pp. 215–252). Edinburgh: Edinburgh University Press.

Laing, L. R. (1969). *Coins and archaeology*. London: Weidenfeld & Nicolson.

Laxton, R. R. (1990). Methods of chronological ordering. In A. Voorrips & B. Ottoway (Eds.), *New tools from mathematical archaeology* (pp. 37–44). Warsaw: Polish Academy of Sciences.

Laxton, R., & Restorick, J. (1989). Seriation by similarity and consistency. In S. P. Q. Rahtz & J. D. Richards (Eds.), *CAA89. Computer applications and quantitative methods in archaeology 1989* (pp. 229–240). Oxford: British Archaeological Reports.

Lipo, C. P., Madsen, M. E., & Dunnell, R. C. (2015). A theoretically-sufficient and computationally-practical technique for deterministic frequency seriation. *PLoS One, 10*(4), e0124942.

Lyman, R. L., & O'Brien, M. J. (2006). *Measuring time with artifacts: A history of methods in American archaeology*. Lincoln, NB: University of Nebraska Press.

McNutt, C. H. (2005). Seriation: Classic problems and multivariate applications. *Southeastern Archaeology, 24*(2), 209–222.

LeBlanc, S. A. (1975). Micro-seriation: A method for fine chronologic differentiation. *American Antiquity, 40*, 22–38.

Madsen, T. (1988). Multivariate statistics and arcaheology. In T. Madsen (Ed.), *Multivariate archaeology: Numerical approaches in Scandinavian archaeology* (Jutland archaeological publications 21) (pp. 7–27). Moesgärd: Jutland Archaeological Society.

Marquardt, W. H. (1978). Advances in archeological seriation. *Advances in Archaeological Method and Theory, 1*, 257–314.

McCafferty, G. (2008). Seriation. In D. M. Pearsall (Ed.), *Encyclopedia of archaeology* (pp. 1976–1978). New York: Academic.

O'Brien, M. J., & Lyman, R. L. (2002). *Seriation, stratigraphy, and index fossils: The backbone of archaeological dating*. New York: Kluwer Academic/Plenum Publishers.

Petrie, W. M. F. (1899). Sequences in prehistoric remains. *Journal of the Anthropological Institute of Great Britain and Ireland, 29*, 295–301.

Phillips, P., Ford, J. A., & Griffin, J. B. (1951). *Archaeological survey in the lower Mississippi Valley 1940-1947*. Cambridge, MA: Peabody Museum of Archaeology and Ethnology.

Pitt-Rivers, A. H. L. F. (1875). On the principles of classification adopted in the arrangement of his anthropological collection, now exhibited at the Bethnal Green Museum. *Journal of the Anthropological Institute of Great Britain and Ireland, 4*, 293–298.

Robinson, W. S. (1951). A method for chronologically ordering archaeological deposits. *American Antiquity, 16*, 293–301.

Scott, A. (1993). A parametric approach to seriation. In J. Andresen, T. Madsen, & I. Scollar (Eds.), *Computing the past: Computer applications and quantitative methods in archaeology, CAA92* (pp. 317–324). Aarhus: Aarhus University Press.

Shennan, S. (1997). *Quantifying archaeology* (2nd ed.). Iowa City: University of Iowa Press.

Wesler, K. W. (1999). Chronological sequences in Nigerian ceramics. *African Archaeological Review, 16*(4), 239–258.

Wilkinson, E. M. (1974). Techniques of data analysis—Seriation theory. *Archaeo-Physika, 5*, 1–142.

19 Stratigraphy

It is not enough to identify layers, although that is, of course, the essential first step; it is the task of the archaeologist to interpret them, to understand the sentence as well as to transliterate it. . . . But it is rarely, in all conscience, that geology deals with us so straight-forwardly; whilst man-made strata are capable of every sort of perversity.

Wheeler (1954: 44–45).

Stratigraphy is the analytical process of studying and interpreting **stratification**, the physical ordering of deposits and interfaces in a stratigraphic series. Thus, as with seriation, it helps us build ordinal scales for chronology.

As noted in Chap. 17, we can distinguish among lithostratigraphy, biostratigraphy, and archaeological or ethnostratigraphy. These share the assumption that we can recognize distinct depositional events and events that remove material (erosion, pit-digging, etc.), but they differ in the unit that constitutes an "event." For a geologist or biostratigrapher, the "event" might last thousands of years. Archaeologically interesting events, by contrast, may have lasted less than a day, or perhaps as long as a couple of centuries.

This chapter concentrates on archaeological stratigraphy, with only occasional reference to its geological cousins. Some authors (e.g., Farrand 1984; Stein 1992) dispute the need for a distinctly archaeological stratigraphy, arguing that lithostratigraphy is sufficient for our needs. Although there are indeed geological analogues for many of the processes that affect archaeological stratification, the goals and shorter time scale of archaeology, as well as the fact that some cultural site-formation processes have no very good geological analogues, arguably calls for a distinctly archaeological stratigraphic theory (Fedele 1984; Harris 1989).

Those on either side of this debate do agree on several points, including the three main principles of lithostratigraphy—the principle of superposition, the principle of original horizontality, and the principle of original continuity—for sedimentary deposits. However, archaeological stratigraphers frequently must interpret deposits and features created by non-sedimentary processes.

The **principle of superposition**: for sedimentary deposits, deposits that were formed earliest underlie ones formed later, so that the ages of deposits are ordered by their sequence of deposition. It is important to keep in mind that, although the deposits accumulated in that order, later events, such as earthquakes, can alter their physical arrangement.

The **principle of original horizontality**: all sedimentary layers formed in bodies of water were originally horizontal, as a result of gravity and other physical phenomena. These layers can later become tilted or warped through tectonic or other forces, however.

The **principle of original continuity**: each layer extended spatially as a whole and uninterrupted sheet or lens, at least until it encountered a barrier, as when sediment accumulates in a trench. Apart from such barriers, any discontinuities or edges that now exist are due to erosion, faulting, animal burrowing, human action or other processes that dislocate or remove portions of the layer.

Archaeologists and geologists can generally agree that an archaeological **deposit** is a three-dimensional "envelope" of sediment distinguishable from other sediments by its physical properties because it formed under particular conditions (Stein 1987: 339; 2001a: 4–5). In Near Eastern archaeology, the deposit or a portion of it is often called a "locus" (Dever and Lance 1978); in other archaeologies, it may be a "layer," "stratum," "facies," "unit," or even "elemental sediment unit" or ESU (Fedele 1976: 34). Stein (1987) argues from the geological perspective that the deposit is the only important stratigraphic unit for both archaeologists and geologists, but

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_19

also notes that they differ in how they define it, the former emphasizing behaviors that created it, and the latter the history of the particles in the deposit (Stein 2001b: 44).

A major difference between the geological and archaeological approaches is in the importance archaeologists assign to the interfaces between deposits, despite the fact that interfaces do have an analogue in geology: the unconformity. Geologists identify unconformities as the upper surfaces of sediments that have been truncated by erosion or lain for a long time without any deposition occurring on them. For archaeologists, an **interface** is a boundary, or surface, between deposits, representing, for example, the surface upon which humans walked around and carried out activities, or the surface created when humans or other agents removed soil or sediment through erosion, excavation, or when they constructed or renovated features. Harris (1989: 43–48) considers an interface to be either the upper surface of non-deposition on a layer, or intrusive features, created by digging, burrowing, gullying or insertion of posts (Vertical Feature Interfaces), or Horizontal Feature Interfaces that result from the destruction or levelling of "upstanding strata," such as stone or brick walls, leaving truncated walls or foundations in place.

Harris (1989) also distinguishes **features**. These have no close analogue in geology and include such non-portable artifacts as hearths, pits, walls, and other structures, many of which have vertical dimensions greater than their horizontal ones and generally have less horizontal extent than layers. Although some features, such as mounds, are equivalent to deposits, others, such as pits, result from the removal, rather than deposition, of material. Of course, pits can be and usually are later filled with deposits, but the pit fill represents one or more depositional events that are distinct from and later than the digging of the pit. Some kinds of features, such as walls, constitute what Harris calls upstanding strata, and would violate the geologists' principle of original horizontality. Every feature is associated with at least one interface, such as the cutting of a pit, or the two sides of a wall.

Finally, it is necessary to mention **arbitrary stratigraphic units**, sometimes called "spits." These have no necessary relation to the "natural" layering of deposits as described above, but are horizontal strips of some arbitrary thickness, such as 5 cm, excavated across an excavation area. These provide an example of grouping artifacts found within the spits by "bounding" (p. 27). It was fairly common for archaeologists to excavate and record artifact context by such arbitrary spits prior to Sir Mortimer Wheeler's (1954) strong advocacy for use of "natural" depositional units. Wheeler convincingly demonstrated that the unevenness of deposits, especially where there were slopes, walls and pit-digging, would lead spits to mix together materials of markedly different ages.

However, arbitrary units have not entirely disappeared from practice. Generally, they are only excusable when a "natural" deposit is apparently so thick that it makes sense, if only out of practicality, to excavate it in increments, which would also act as a check in case it turned out that there were significant differences from top to bottom that field personnel had failed to recognize on the basis of lithological characteristics. It is particularly misleading to use horizontal spits as stratigraphic units on sloping sites, such as the talus slopes of rockshelters, although this practice unfortunately persists in some quarters (Ward et al. 2016). Unless they are completely contained within a deposit, spits make a poor basis for any kind of chronological analysis (low validity).

19.1 Traditional Stratigraphic Analysis

Archaeologists have employed stratigraphy informally as early as Low's (1775) cross-section of a Scottish *tumulus*, but it took a long time before it became an explicit archaeological practice.

The key tool for archaeological stratigraphy prior to about 1980 was the **stratigraphic section** or profile. This, somewhat like the geologist's or pedologist's profile, but emphasizing archaeological deposits and features over lithological ones, depicts a vertical slice through a portion of an archaeological site. Early versions were schematic, sometimes reconstructed after excavation was complete to demonstrate, for example, how tumuli were constructed (Harris 1989: 50). However, some fairly early ones appear to depict stratification in the side of a trench (e.g., Fig. 19.1; von Luschan 1893: 119), and Darwin (1882) even mentions that he drew his sections by measuring from a taut, horizontal string, as field archaeologists often do even today.

Modern sections most commonly are literal depictions of one or more "walls" of an excavation area, or a "baulk" of sediment left standing with excavation on either side, after fieldworkers trimmed these walls to be very nearly vertical (Fig. 19.2), where this would not pose a safety hazard. While the deposits visible in such a section are relatively moist, it is often easy to see color differences that allow us to recognize interfaces between deposits quite well. In some instances, such differences may be subtler, requiring more skill to identify. Excavators did, and still do, draw these sections as one of their main documents of the stratification through which they excavated. Although archaeologists also photograph such sections, drawing is indispensable, as the resulting profile constitutes the archaeologist's interpretation of the stratification, while a photograph only represents certain aspects of its visual characteristics, and evidence that the archaeologist can see and even feel might not show in a photograph.

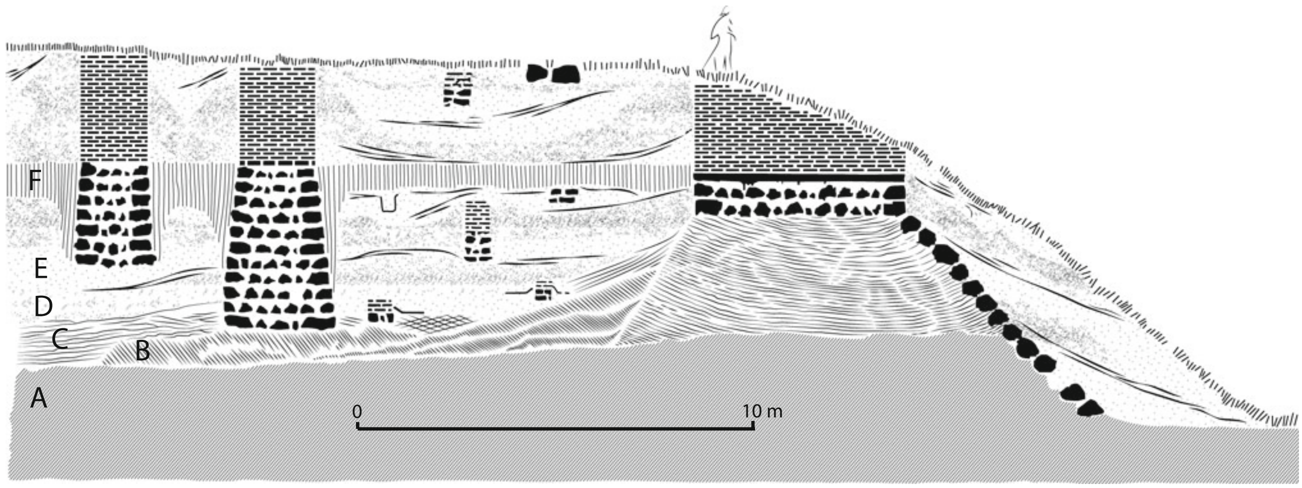


Fig. 19.1 Section through part of the tell site of Zinjirli near the Syrian-Turkish border during German excavations in the early 1890s, with major strata labelled A-F. (Y. Salama, after von Luschan 1893)

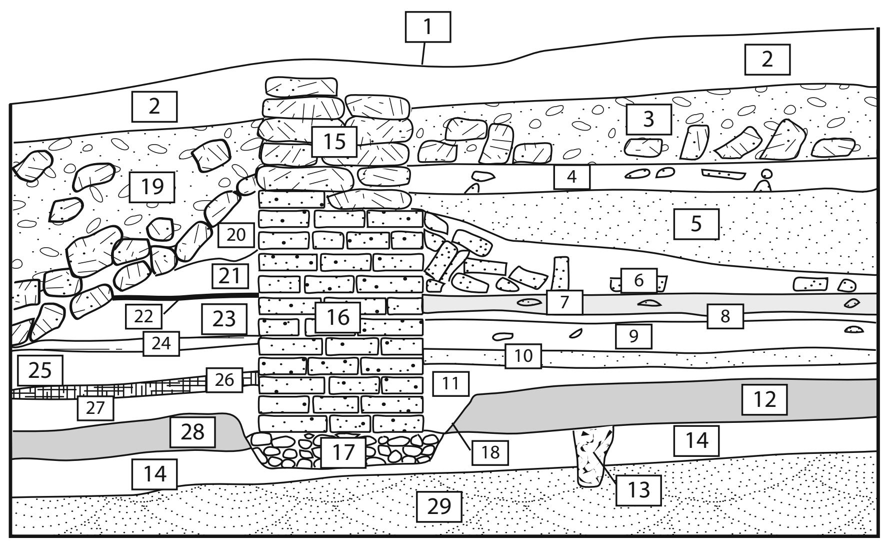


Fig. 19.2 Example of a stratigraphic section drawing to show the vertical relationships among deposits in the plane of one edge of an excavated 3 m × 3 m excavation unit. (Modified from Wheeler 1954: fig. 16a)

One disadvantage of the traditional section is that it only accounts for deposits, features and interfaces that happen to intersect the vertical plane of the section, usually the four walls of a rectangular excavation area. Some field archaeologists attempt to compensate for this problem by adding supplementary sections through important features, such as pits or ovens, that do not intersect any of the side walls. However, "open-area excavations" do not employ any baulks and, except at the edges of whole excavation areas, can only "reconstruct" sections from heights taken by total station or GPS at multiple locations on the surfaces of layers and features.

Another traditional tool of the archaeological stratigrapher is a chart that summarizes the stratigraphy of one or more sites, including associations between unconnected stratigraphic series (see multilinear sequence, below), and often estimated dates or significant artifact content (Fig. 19.3). Charts such as these have been a key component of regional chronologies for many decades, but especially prior to the availability of radiocarbon dating.

Today, largely thanks to the fact that geologists also depict stratification, there is software that allows us not only to depict two-dimensional views of vertical sections, but to construct three-dimensional models of stratification, whether based on data from contiguous or non-contiguous excavations and exposures, or from cores and bore-holes (Fig. 19.4). Among the software that helps us accomplish this and other kinds of geological visualizations are GEO5, MatStrat, Strater, and Rockworks. Archaeologists can also create three-dimensional models of excavated stratification in a Geographic Information System (GIS).

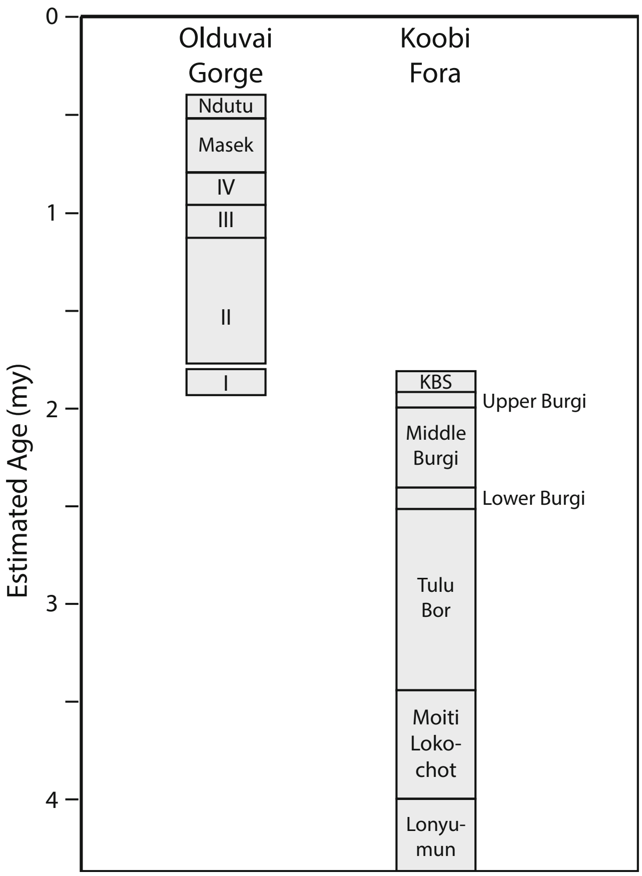


Fig. 19.3 Chart to summarize and compare the major stratigraphy at Olduvai Gorge and Koobi Fora. (Dates of transitions estimated from Stanistreet 2012; Gathago et al. 2008)

19.2 The Harris Matrix

A key tool for modern archaeological stratigraphy is the **Harris matrix**, an abstract representation built from the unequivocal relationships between layers, interfaces and features (Harris 1975). It is a bit like a wiring diagram, flow chart, or lattice, and not really a matrix in the mathematical sense (Orton 1980), leading some to prefer the term, sequence diagram. Because it does not depend on sections, it can be used in open-area excavations as well as ones that employ baulks.

A Harris matrix depicts each layer, feature and interface and the non-redundant stratigraphic relationships among them by line segments (Fig. 19.5). Its purpose is not to show physical relationships, but rather the sequence of deposits, features and interfaces in time (Harris 1989: 34–36).

The "boxes" in the diagram are features, deposits or interfaces that represent a distinct "event" in time, no matter its duration. Vertical line segments connect boxes in accordance with the **Law of Stratigraphic Succession**: any stratigraphic unit in the diagram belongs "between the undermost (or earliest) of the units [that] lie above it and the uppermost (or latest) of all the units [that] lie below it and with which the unit has some physical contact, all other superpositional relationships being redundant" (Harris 1989: 157–158). There may be, and often are, many other physical and chronological relationships that we omit from the diagram because they would not contribute any further *chronological* information and would make the diagram much too

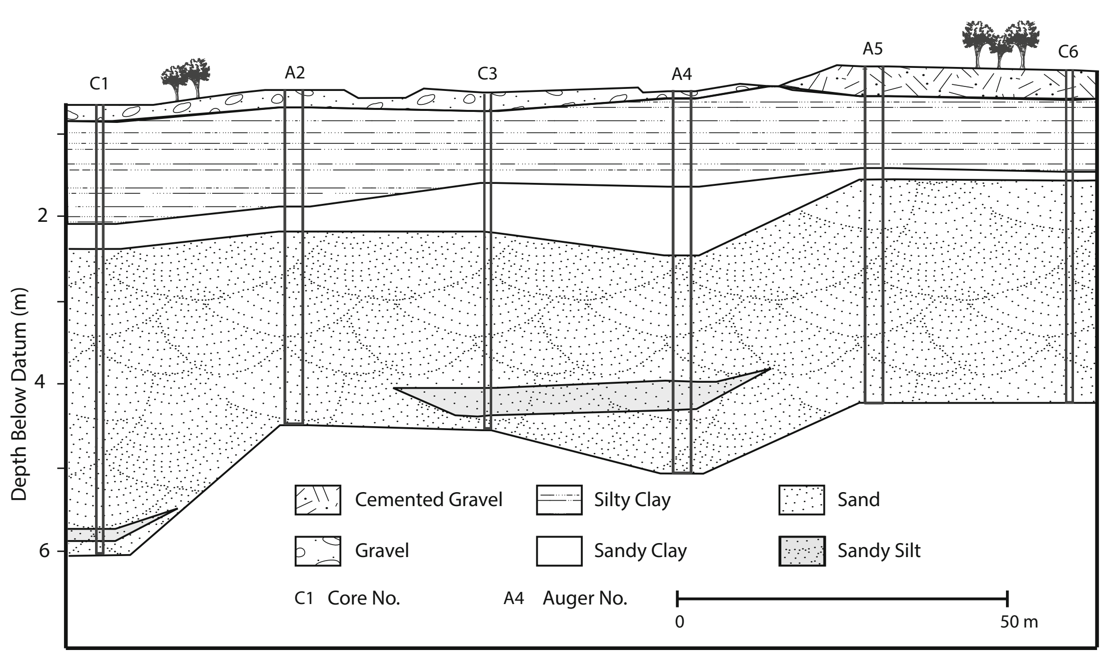


Fig. 19.4 Stratification over part of a landscape reconstructed from information in a transect of cores (narrow rectangles) and auger holes (wider rectangles)

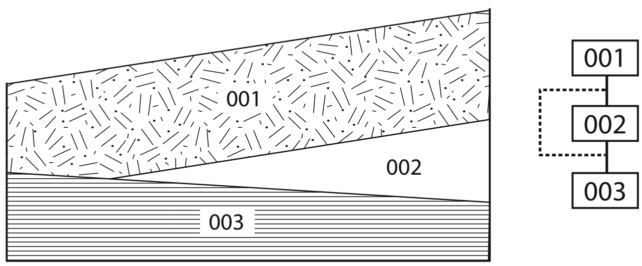


Fig. 19.5 A very simple stratigraphic sequence of three layers and its accompanying Harris matrix. Note that the relationships for the physical superposition of 001 over 003 (dashed line segments) is redundant, and so should be removed, as Harris matrices only show the chronological relationships, not the physical ones

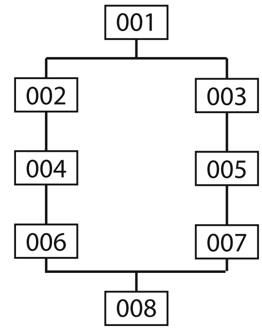


Fig. 19.6 Multilinearity in a stratigraphic sequence caused by a wall (008) that separates the deposits on either side. Although the sequence in each of the two branches is clear, there is no valid way to determine the relationships between stratigraphic contexts in different branches. While it is possible that layer 004 was deposited at the same time as layer 005, it might just as well be older or younger than 005

complicated (Harris 1989: 112). The oldest units occur at the bottom of the diagram and the youngest at the top.

We can also represent stratigraphic relationships and the issue of superpositional redundancy mathematically, taking advantage of stratigraphy's deductive nature. If we use the ">" sign to mean "older than," we can make statements such as,

Unit 1 > Unit 2, and
Unit 2 > Unit 3

From this, we can deduce that Unit 1 > Unit 3, so we do not need to show line segments in the diagram for the Unit 1 > Unit 3 relationship. In fact, we know that Unit 2 intervenes between Units 1 and 3, and so can infer that there was some passage of time, however brief or long, between the two events that Units 1 and 3 represent. In other words, the Harris matrix should only show as many relationships as are necessary to represent the sequence of such events. Also, should we know that Unit 2 = Unit 4 (e.g., excavators in two adjacent excavation areas applied two different numbers to what was clearly the same deposit, or two walls numbered 2 and 4 were clearly bonded together in such a way as to show that they were built simultaneously), we could deduce that Unit 4 is also older than Unit 1 and younger than Unit 3.

The resulting diagram is an extremely useful representation of the sequence of all depositional, construction and removal events, including ones whose traces (deposits, features or interfaces) do not intersect with any stratigraphic section or profile. Thus, it is superior to the traditional profile in its ability to represent the relative chronology.

However, it is also the case that some stratigraphic relationships are uncertain because there is no physical contact between stratigraphic units. For example, in Fig. 19.6, it is clear that Units 004 and 005 are both older than Unit 001 and younger than Unit 008, but we have no stratigraphic information that would allow us to confirm or deny that Unit 004 > Unit 005. Mathematically, we could represent these relationships as.

Unit 008 > Unit 006 > Unit 004 > Unit 002 > Unit 001, and
Unit 008 > Unit 007 > Unit 005 > Unit 003 > Unit 001

There is no way for deduction to help us decide whether Unit 004 is older, younger, or contemporary with Unit 005. This constitutes what Harris calls a multilinear stratigraphic sequence, and such sequences are particularly common on sites that have substantial architecture because ditches, walls, and other constructed features commonly partition a site's deposits into a patchwork of regions with separate stratigraphic histories. It also happens when we have non-adjacent excavation areas or adjacent areas that became narrower with depth of excavation. Such cases demand non-stratigraphic information, such as radiocarbon dates (Chap. 20) or seriation of artifact content (Chap. 18), to help us decide the most probable associations between unconnected units (e.g., Triggs 1993). In the absence of such evidence, it often happens that a particular context could "slide" up and down a vertical line in the Harris matrix (Fig. 19.6).

Despite the problem of multilinear sequences, archaeologists typically try to "phase" their Harris matrices by estimating the most likely positions of such "sliding" units or by grouping them together. Often, this is on the basis of artifact content or sediment characteristics, sometimes on the basis of other dating evidence, such as historic maps. However, all such estimates will be preliminary and prone to varying degrees of error.

Some archaeologists have augmented the original Harris matrix by trying to indicate the estimated lifespans of stratigraphic units by elongating certain boxes vertically, or by symbolically distinguishing layers, interfaces and various kinds of features to make the diagram easier to interpret (e.g., Fig. 19.7; Bibby 1993; Dye and Buck 2015; Hammond 1993; Paice 1991). Some of these depart from Harris's main goal of explicating the purely chronological sequence of events, however.

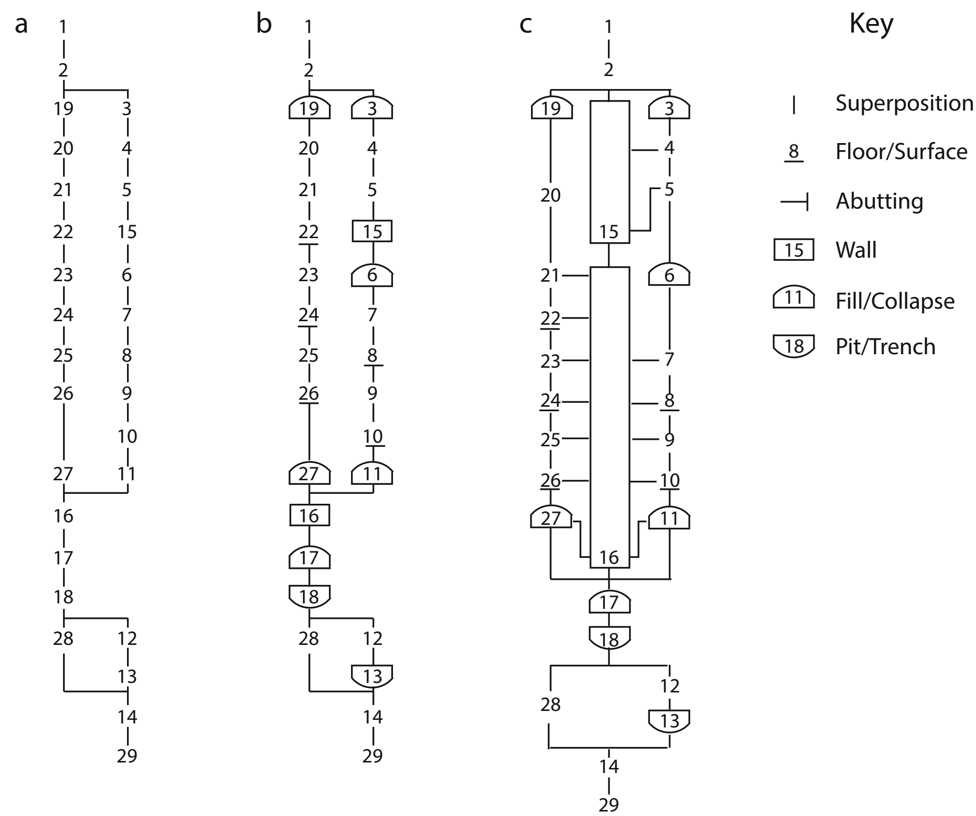


Fig. 19.7 Conventional Harris matrix (**a**) for the section in Fig. 17.2 (after Paice 1991), a modified version (**b**) that uses different shapes to distinguish walls from pits, layers, and other types of contexts, and another version (**c**) that enlarges units vertically to indicate the longevity of certain walls and features

In addition, some archaeologists omit most interfaces from Harris matrices when they are only boundaries between deposits, rather than so-called living surfaces or the result of digging pits or trenches. In the case of relatively long-lived interfaces that archaeologists often describe as "surfaces" or "floors," it is possible that artifacts and other material found right at the interface is more chronologically and culturally important than the artifacts found in the underlying deposit, as they may have resulted from single or repeated episodes of use, discard, and site maintenance, rather than being imported in construction fill. Consequently it makes sense to give such interfaces their own identifying labels.

Constructing a Harris matrix is a pretty simple task when there are only a few units to represent, but they can quickly become very complicated, particularly for large or stratigraphically complex sites. That is one of their advantages—the ability to condense a lot of complex stratigraphic information—but it also makes their manual construction more difficult, and may lead to such obvious errors as "crossovers" in the superpositional relationships (Triggs 1993: 252, 256). Thankfully, there are software packages that help with this. Users enter all the relevant individual stratigraphic relationships, such as Unit 1 > Unit 2 > Unit 3, and the software employs deductive logic to work out the whole matrix. In addition, it will flag any relationships that are "impossible" by deduction. For example, errors in recording or in the field might yield such relationships as Unit 1 > Unit 2 > Unit 3 > Unit 1, a clearly false statement that the software would identify. It is then incumbent on the archaeologist to look into the field notes to see where the error lies and correct it. Often, the problem is that an excavator incorrectly gave two distinct deposits the same number just because they had similar lithological characteristics. At time of writing, some of the software packages available for stratigraphic analysis are ArkMatrix, Harris Matrix Composer, Stratify, and Strati5, while ArchEd is older software that draws Harris matrices without analysis (Sikora et al. 2016).

19.3 Single-Context or Single-Layer Plans

Although Harris is best known for the Harris matrix, he also argues strongly for the use of single-layer plans (1989: 73–79). These are particularly important for deposits or features that do not intersect any of the available sections, not to mention open-area excavations that do not employ standing sections. They consist of layered maps, each showing the spatial extent of only one deposit, feature or interface (Fig. 19.8), so that stacking them would, in a sense, create a highly simplified three-dimensional model of the stratification.

Today, it would be possible to create these single-context plans in a GIS in which each is a different layer. The GIS also allows us to make them more realistic by giving the various

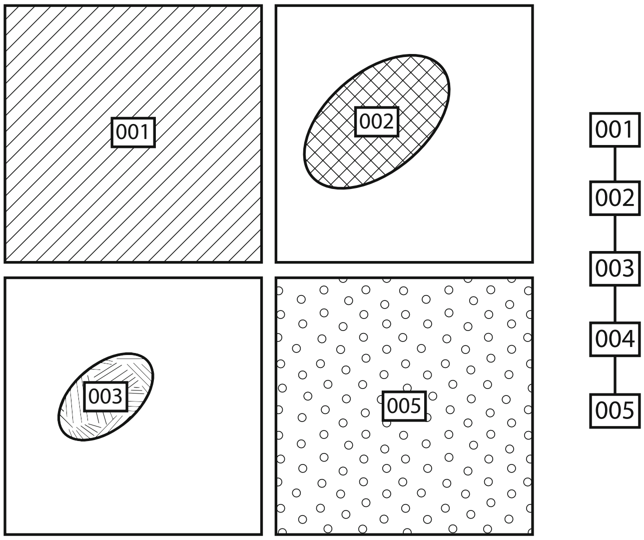


Fig. 19.8 Examples of single-context plans. Unit 005 is a layer into which a pit (interface 004) was cut, then filled by deposits 003 and 002, before the whole area was covered by layer 001. The single-context plan for interface 004 would look the same as that for pit fill 002, but with contour lines inside it

stratigraphic units realistic thickness or depth, and including variation in that thickness, as determined from top and bottom levels that fieldworkers recorded with theodolites, total stations, differential GPS, or photogrammetry (e.g., Neubauer 2004: 162–164).

19.4 Grouping Stratigraphic Units

Once archaeologists have understood a site's stratification, nowadays typically with a Harris matrix, they usually group units in such a way that ones that are very nearly contemporary with one another, or that belong to a particular period of site occupation, are associated. This usually results in a larger unit than an individual deposit that Old World archaeologists often call a "phase" (not to be confused with Willey and Phillips's [1956] usage of the term) or "stratum," and that New World archaeologists might call a "component." Stein (1992) would call this concept an "ethnozone." The phase, component, or ethnozone then becomes one of the building-blocks for making comparisons with other sites, creating regional sequences such as those in the traditional stratigraphies (e.g., Fig. 19.3), and for inferring the regional extent of "cultures" or "complexes."

We also attempt to associate units from non-contiguous excavation areas, or even from different sites. Geologists, prior to radiometric dating, depended heavily on "type fossils" or on sets of animal remains or pollen to establish whether spatially separated deposits were probably contemporary. Their problem was analogous to the archaeological problem of multilinear sequences. Recently, like archaeologists, they use multivariate analysis of fossil content and other physical attributes for this purpose (Birks 1987).

Phasing, whether it involves grouping units in a superpositional sequence in one location or ones that are spatially separated, relies in part on non-stratigraphic information. As with paleontologists' and geologists' use of type fossils, this information is typically the artifact content of deposits. We can analyze the distributions of artifact types in deposits, using many of the same methods discussed in Chaps. 3, 6 and 18 (Triggs 1993), in order to group deposits whose content is statistically likely to come from the same population.

19.5 Confounding Factors

While such methods are useful, we cannot assume that all artifacts and "ecofacts" found in a deposit or lying on an interface are of the same age as the deposit or interface itself. Natural and cultural site-formation processes (Schiffer 1987) can sometimes introduce materials to deposits that differ in age from the deposit, while the fact that archaeological surveys routinely find ancient artifacts on modern surfaces demonstrates that deposits and interfaces can be palimpsests of materials that vary in age. It is useful to distinguish three kinds of remains of varying relationship to the deposit in which they are found (Harris 1989: 93; Wood and Johnson 1978; see also Chap. 20).

Indigenous remains are those artifacts and ecofacts that were created only shortly before the deposit in which they were found. These are the remains that we ideally want to use for defining phases or components.

Residual remains are artifacts or ecofacts that are significantly older than the deposits in which we find them. They include curated artifacts, such as family heirlooms and certain coins, that were saved or circulated for many decades after their manufacture before eventually being lost or intentionally buried. Most commonly, they include seeds, charcoal, artifacts and other items that were removed from some other sediment by pit-digging, erosion, plowing, or some other process and redeposited in a "higher" stratigraphic position.

Infiltrated remains are artifacts and ecofacts that were created after the deposit in which they were found, but somehow worked their way into that deposit without leaving any obvious trace of their infiltration. Among the processes that can cause younger objects to settle into older deposits are earthworm activity, frost heaving, and other natural site-formation processes that cause size-sorting of particles.

19.6 Unitary Association Method

One relatively new method for grouping deposits to create a regional stratigraphic sequence is based on set theory. Geologists developed the Unitary Association Method (UAM) to establish stratigraphic relationships for non-contiguous deposits on the basis of sets of fossils present in them, and on the assumption that each type of fossil has a chronological “range zone” (Fig. 19.9; Guex 1991; Guex and Galster 2016; Guex et al. 2016; Savary and Guex 1991). Blackham (1998) adapted UAM for use with archaeological assemblages.

Unlike seriation, UAM uses the observed superpositions of artifact (or fossil) taxa in stratified series to identify the sequence of associated sets of artifacts as well as “virtual associations” that take into account the fact that any sampled deposit is unlikely to include all the artifact types that existed at the time it formed (Fig. 19.10; recall bias in diversity measures, pp. 124-125). A “local horizon” is the set of all associated artifact types in a layer, but those that are sub-sets of other local horizons are combined with them to create “maximal horizons.” Other artifacts associated with the members of each maximal horizon are used to define “maxi-

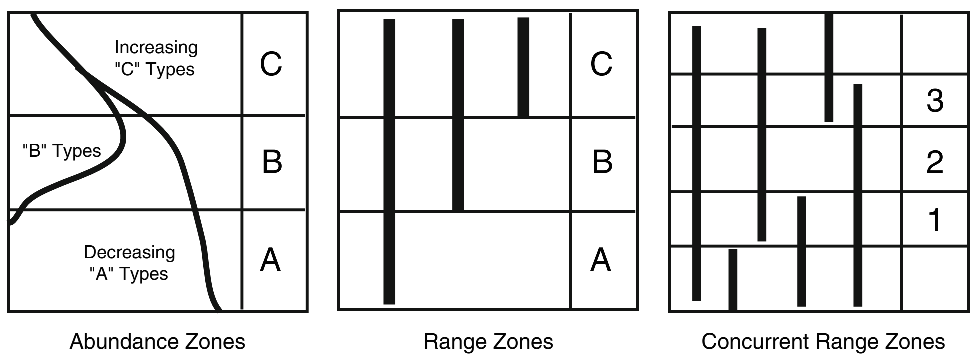


Fig. 19.9 Three different models for the co-occurrence of artifact types over time, (**a**) frequency seriation, (**b**) range zones that identify the earliest and latest known occurrence of each type, and (**c**) concurrent range zones. (After Guex 1991)

Fig 19.10 Summary of processes for linking stratigraphy across three hypothetical sites by UAM. Artifact types 1–10 found in each layer at three sites are grouped into unique Local Maximal Horizons (**a**), and redundancies combined to create Residual Maximal Horizons (**b**). Artifacts from the “neighborhoods” of these are added to create Maximal Cliques (**c**), which are then reordered as “Petrie” or “P-matrix” (**d**, see Chap. 18). Combining MCs to remove redundancy creates Unitary Associations (**e**), which can then be used to determine which layers from the three sites are contemporary (**f**, after Blackham 1998). Note that “3-3” means that the layer only belongs to UA 3, while “1-2” means that the layer (in this case 1.1) could belong to either UA 1 or UA 2

a

Site	LMH	Artifacts
1	5	1, 2, 8
	4	4, 5, 8
	3	7, 8
	2	7, 9, 10
2	4	1, 2, 3
	2	5, 6
	1	9, 10
3	4	4, 6
	2	6, 7, 8, 9
	1	10

b

RMH	Site. Layer	Artifacts
1	3.2	6, 7, 8, 9
2	2.4	1, 2, 3
3	1.5	1, 2, 8
4	1.4	4, 5, 8
5	1.2	7, 9, 10
6	3.4	4, 6
7	2.2	5, 6

c

MC	Site. Layer	Artifacts
1	1.4	4, 5, 6, 8
2	3.2	6, 7, 8, 9
3	1.5	1, 2, 8
4	2.4	1, 2, 3
5	1.2	7, 9, 10

d

MC	Site. Layer	Artifacts
4	2.4	1, 2, 3
1	1.5	1, 2, 8
2	1.4	4, 5, 6, 8
5	3.2	6, 7, 8, 9
	1.2	7, 9, 10

e

UA	Artifacts
4	1, 2, 3, 8
3	4, 5, 6, 8
2	6, 7, 8, 9
1	7, 9, 10

f

Site 1		Site 2		Site 3	
Layer	UA	Layer	UA	Layer	UA
5	4-4	4	4-4		
		3	4-4		
4	3-3	2	3-3	4	3-3
3	2-2			3	2-2
				2	2-2
2	1-1	1	1-1	1	1-1
1	1-2				

mal cliques," which are ethnostratigraphic (or, for paleontologists, biostratigraphic) units that represent unique associations of artifact types over all the sites included in the analysis. Information on the superposition of artifacts that belong to these associations allows us to put the maximal cliques into correct stratigraphic order. Software (Savary and Guex 1991, 1999; Guex et al. 2016) resolves deductive contradictions in the ordering and produces an ordered series of "unitary associations." Subsequently, the contents of each physical stratigraphic unit at each site must be a subset of one of the unitary associations, and so we can order them, even relative to deposits with which they have no physical connection or, less intuitively, even no artifacts in common.

Although UAM is used in paleontology for biochronology (e.g., Galster et al. 2010), it remains uncommon in archaeology, but has potential to make very useful contributions to regional chronologies (Blackham 2002).

19.7 Summary

- Superposition of deposits, features and interfaces at stratified sites provides physical evidence for an ordinal chronological scale
- Stratigraphy, or interpreting stratification, is a primarily deductive process, but also depends on archaeologists' careful and informed observations, and their documentation in section drawings, photographs, single-context plans, and Harris matrices
- The Harris matrix is a convenient and very useful tool for discerning, interpreting and displaying the sequence of events that resulted in a site's stratification
- Other tools allow us to "phase" the stratigraphic units both within sites and across regions
- The Unitary Association Method (UAM) allows us to extend stratigraphic sequences across regions on the basis of "index fossils" with known sequences

References Cited

Bibby, D. I. (1993). Building stratigraphic sequences on excavations: An example from Konstanz, Germany. In E. C. Harris, M. R. Brown, & G. J. Brown (Eds.), *Principles of archaeological stratigraphy* (pp. 104–121). London: Academic.

Birks, H. J. B. (1987). Multivariate analysis of stratigraphic data in geology: A review. *Chemomatics and Intelligent Laboratory Systems, 2*(1–3), 109–125.

Blackham, M. (1998). The unitary association method of relative dating and its application to archaeological data. *Journal of Archaeological Method and Theory, 5*, 165–207.

Blackham, M. (2002). *Modeling time and transition in prehistory: The Jordan valley chalcolithic (5500–3500 BC)*. Oxford: British Archeological Reports.

Darwin, C. (1882). *The formation of vegetable mould, through the action of worms with observations on their habits*. London: John Murray.

Dever, W., & Lance, H. D. (1978). *A manual of field excavation*. New York: Hebrew Union College.

Dye, T. S., & Buck, C. E. (2015). Archaeological sequence diagrams and Bayesian chronological models. *Journal of Archaeological Science, 63*, 84–93.

Farrand, W. (1984). Stratigraphic classification: Living within the law. *Quarterly Review of Archaeology, 5*, 1.

Fedele, F. G. (1976). Sediments as palaeo-land segments: The excavation. In D. A. Davidson & M. L. Shackley (Eds.), *Geoarchaeology: Earth science and the past* (pp. 23–48). London: Duckworth.

Fedele, F. G. (1984). Towards an analytical stratigraphy: Stratigraphic reasoning and excavation. *Stratigraphica Archaeologica, 1*, 7–15.

Galster, F., Guex, J., & Hammer, Ø. (2010). Neogene biochronology of Antarctic diatoms, a comparison between two quantitative approaches, CONOP and UAgraph. *Palaeogeography, Palaeoclimatoloty, Palaeoecology, 285*, 237–247.

Gathago, P. N., Brown, F. H., & Mcdougall, I. (2008). Stratigraphy of the Koobi Fora formation (Pliocene and Pleistocene) in the Loiyangalani region of Northern Kenya. *Journal of African Earth Sciences, 51*, 277–297.

Guex, J. (1991). *Biochronological correlations*. Berlin: Springer.

Guex, J., & Galster, F. (2016). A simple technique to establish sequences of datums and to highlight transgressive-regressive cycles. *Carnets de Géologie, 16*, 1–16.

Guex, J., Galster, F., & Hammer, Ø. (2016). *Discrete biochronological time scales*. New York: Springer.

Hammond, N. (1993). Matrices and Maya archaeology. In E. C. Harris, M. Brown, & G. Brown (Eds.), *Principles of archaeological stratigraphy* (pp. 139–154). London: Academic.

Harris, E. C. (1975). The stratigraphic sequence: A question of time. *World Archaeology, 7*, 109–121.

Harris, E. C. (1989). *Principles of archaeological stratigraphy* (2nd ed.). London: Academic.

Low, G. (1775). Account of a tumulus in Scotland. *Archaeologia, 3*, 276–277.

Neubauer, W. (2004). GIS in archaeology—The interface between prospection and excavation. *Achaeological Prospection, 11*, 159–166.

Orton, C. (1980). *Mathematics in archaeology*. London: Collins.

Paice, P. (1991). Extensions to the Harris matrix system to illustrate stratigraphic discussion of an archaeological site. *Journal of Field Archaeology, 18*, 17–28.

Savary, J., & Guex, J. (1991). BioGraph: Un nouveau programme de construction des correlations biochronologiques basées sur les Associations Unitaires. *Bulletin Laboratoire Géologie, Université de Laussanne, 313*, 317–334.

Savary, J., & Guex, J. (1999). Discrete biochronological scales and unitary associations. Description of the BioGraph computer program. *Mémoires de Géologie, Lausanne, 34*, 1–281.

Schiffer, M. B. (1987). *Formation processes of the archaeological record*. Albuquerque: University of New Mexico Press.

Sikora, J., Stroka, J., & Tyszkiewicz, J. (2016). Strati5—Open mobile software for Harris matrix. In S. Campana, R. Scopigno, G. Carpentiero, & M. Cirillo (Eds.), *CAA2015: Keep the revolution going* (Vol. 2, pp. 1005–1014). Siena: Università di Siena.

Stanistreet, I. G. (2012). Fine resolution of early hominin time, Beds I and II, Olduvai Gorge, Tanzania. *Journal of Human Evolution, 63*, 300–308.

Stein, J. K. (1987). Deposits for archaeologists. *Advances in Archaeological Method and Theory, 11*, 337–393.

Stein, J. K. (1992). Interpreting stratification of a shell midden. In J. Stein (Ed.), *Deciphering a shell midden* (pp. 71–93). San Diego: Academic.

Stein, J. K. (2001a). Archaeological sediments in cultural contexts. In J. K. Stein & W. R. Farrand (Eds.), *Sediments in archaeological context* (pp. 1–28). Salt Lake City: University of Utah Press.

Stein, J. K. (2001b). A review of site formation processes and their relevance to geoarchaeology. In P. Goldberg, V. T. Holliday, & C. R. Ferring (Eds.), *Earth sciences and archaeology* (pp. 37–51). New York: Kluwer Academic/Plenum Publishers.

Triggs, J. (1993). The seriation of multilinear stratigraphic sequences. In E. C. Harris, M. Brown, & G. Brown (Eds.), *Practices of archaeological stratigraphy* (pp. 250–273). London: Academic.

von Luschan, F. (1893). *Ausgrabungen in Sendschirli* (Vol. 1). Berlin: W. Spemann.

Ward, I., Winter, S., & Dotte-Sarout, E. (2016). The lost art of stratigraphy? A consideration of excavation strategies in Australian indigenous archaeology. *Australian Archaeology, 82*(3), 263–274.

Wheeler, R. E. M. (1954). *Archaeology from the earth*. Oxford: Oxford University Press.

Willey, G. R., & Phillips, P. (1956). *Method and theory in American archaeology*. Chicago: University of Chicago Press.

Wood, W. R., & Johnson, D. L. (1978). A survey of disturbance processes in archaeological site formation. *Advances in Archaeological Method and Theory, 1*, 315–381.

Chronometric Dating

20

... there is still a state of anarchy surrounding publication and use of dates. The dates themselves are quoted on this or that half-life, without a statement of which is being used and now ... we have the added confusion of recalibration. Partly the problem is the failure of prehistorians to understand the basic principles and statistical nature of C14 dating, and partly due to their blatantly ignoring the advice given by the physicists....

Collis (1971: 200).

Sometimes archaeologists have to satisfy themselves with the ordinal time measures that stratigraphy, typology or seriation can provide to study change over time but, whenever possible, they prefer to date deposits and other events on a time scale in years. In other words, they like to have an interval scale for time.

This chapter reviews the principles and challenges involved in what many archaeologists call "chronometric dating" or sometimes "absolute dating," with emphasis on two dating methods, radiocarbon dating and dendrochronology. However, most of the principles and problems are ones that all chronometric methods share.

This is not the place to discuss the physics or environmental and biological processes that lie behind some of these methods, or the details of sample preparation and instrumental measurement, for which there are many good overviews (e.g., Aitken 1990; Bowman 1990; Brothwell and Pollard 2008; Malainey 2011: 91–168; Speer 2010; Taylor and Aitken 1997; Taylor and Bar-Yosef 2014). The chapter's focus will be on how we interpret and apply the results of these methods.

20.1 Dates and Events

Before using any chronometric method, it is critical to specify the event or events you are trying to date, and how accurate and precise the dates should be if they are to be of real value. Archaeologists are sometimes too vague about this. Saying that you want "to date the site" is simply not sufficient, and a haphazard collection of radiocarbon or any other kind of dating evidence from a site does not magically result in anything meaningful.

Instead, it is best to refer to specific events, such as the initial foundation of the site, or the site's abandonment, or a particular construction or destruction event. Sometimes we have historical information that allows us to date such events very closely, such as the destruction of Pompeii by Mount Vesuvius's eruption in A.D. 79. This can lead to what we sometimes describe as a "point estimate" of the date, as in Fig. 20.1a, even though it is really only a "point" in the sense of having high precision.

Other kinds of events that could be of interest include the death of an individual whose skeleton is in a grave, the digging of the grave-pit, or the construction, renovation, or demolition of a house. Most of the time, we cannot make reasonable point estimates for the dates of such events, but instead date them to an interval, like "fourth century BCE" or "8500–7800 years ago." Since we often want to use chronometric methods to date the ordinal changes found through stratigraphy or seriation, our event of interest can also be the deposit of a particular layer or the transition from the Early to Middle Woodland Period.

It is particularly important to remember that the event that our dating method dates is not necessarily the event in which we are most interested. Consequently, we are dealing with indirect measurement, for which we must defend the validity. The event of the formation of a tree ring, for example, is at best indirectly related to archaeological events, like the construction of a house or the last use of a hearth. Some archaeologists

Electronic Supplementary Material: The online version of this chapter (https://doi.org/10.1007/978-3-030-47992-3_20) contains supplementary material, which is available to authorized users.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_20

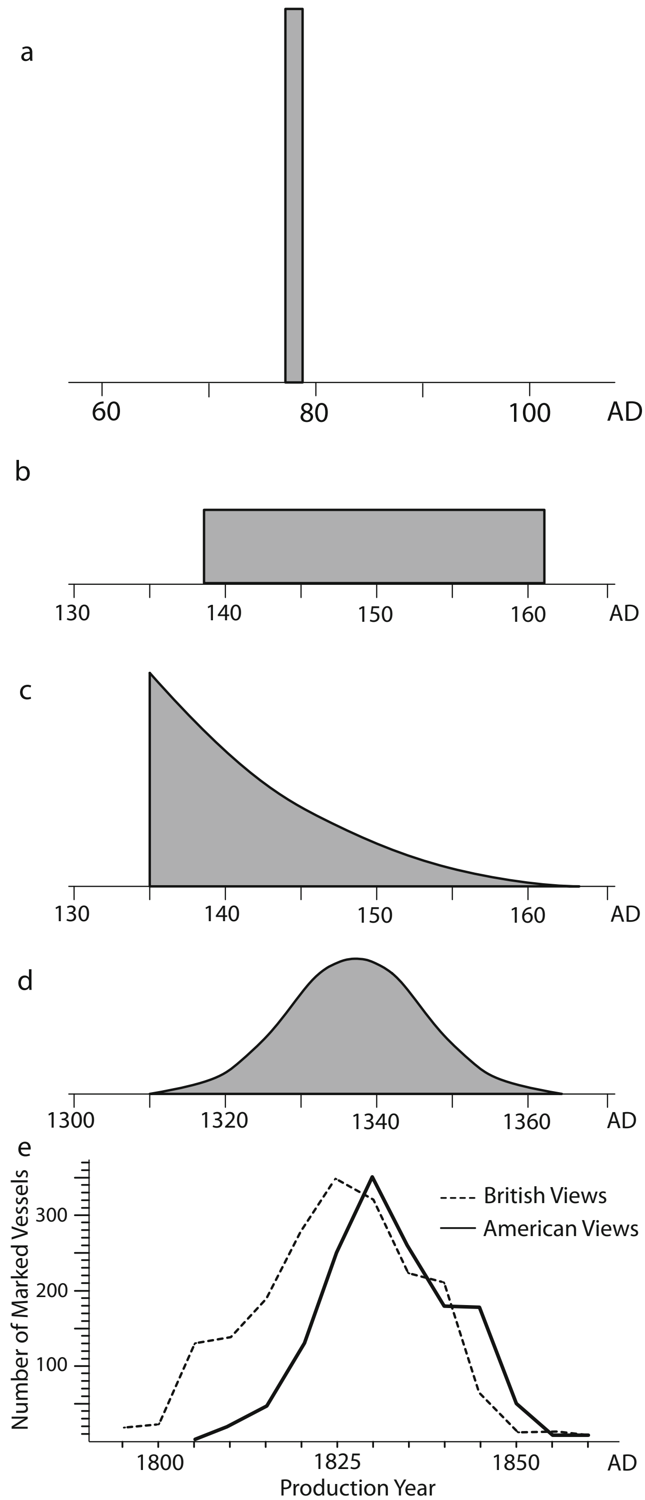


Fig. 20.1 Probability density models for different kinds of date: (**a**) a point estimate of AD 79, (**b**) a uniform distribution for the rule of Antoninus Pius, a date range of AD 138–161, (**c**) an exponential curve to model a *terminus post quem* of AD 135, (**d**) a Gaussian distribution to indicate a date of AD 1336 ± 7, and (**e**) a more informative model based on the distribution of Staffordshire pottery with views of British and American landmarks, dated by makers' marks. (Y. Salama, after Samford 1997 and cf. Orton 1980: 100)

refer to these differences as "offsets," sometimes more specifically defined as an age difference caused by "reservoir effects" in the radiocarbon method (see "Sources of Error in Radiocarbon Dating") or as the difference between a tree's cutting date and the date of the outermost preserved tree ring. "Offset" is, however, a rather imprecise term.

Quite often, the event that we are able to date precedes the event in which we are interested, and the traditional term for the date on such an event is ***terminus post quem*** (TPQ). For example, if we find a coin in an undisturbed archaeological deposit, and we know that the coin was made in AD 1701, that tells us that the deposit could not have formed any earlier than 1701.

Conversely, we sometimes have a date that must be later than the event of interest. The technical term for that is ***terminus ante quem*** (TAQ). For example, if we are trying to date the digging of a pit, and the pit is stratigraphically below a building that we know, either from documentary records or a dated cornerstone, was built in 1744, the pit must have been dug before the building's construction. Consequently, the event cannot have been any later than 1744 and could have occurred much earlier. The trouble with TPQ and TAQ is that they only constrain our dating knowledge at one end of a potentially large range.

One way to model these situations is to use a probability density function (see Chap. 8). We might model a point estimate, as in Fig. 20.1a or a specific time range for a pottery type as in Fig. 20.1b. We could represent a TPQ very simply, with a uniform distribution that extends from the TPQ all the way to the present, indicating no confidence in narrowing down the date. Alternatively, if we think the event of interest likely only followed the *terminus post quem* by a short period, with later dates having less and less probability, we might assume a model of exponential decrease in that probability, as in Fig. 20.1c. There might also be reason to suspect some delay between the TPQ and the event of interest, in which case the modelled curve might rise for a few years to the right of the TPQ before decreasing exponentially, instead of decreasing immediately. The shaded area in each graph represents our estimated probability that the event of interest occurred in a particular date range.

Dean (1978) offers a very useful classification of what he terms dating discrepancies that offers clear alternatives to the somewhat vague term, "offset." He distinguishes between the **target event** (the event of interest), the **reference event** (the potentially datable event that we think is close to the target event), and the **dated event** (the event that we are actually able to date). If we are really lucky, all of these are equivalent, but there are typically differences between them. If so, using a reference event as a proxy for a target event will lead to bias if we do not take that difference into account. Dean calls errors that result from the reference event being earlier than the target event, dating **disjunctions**, which are related to the concept of *termini post quem*. He calls ones that result from

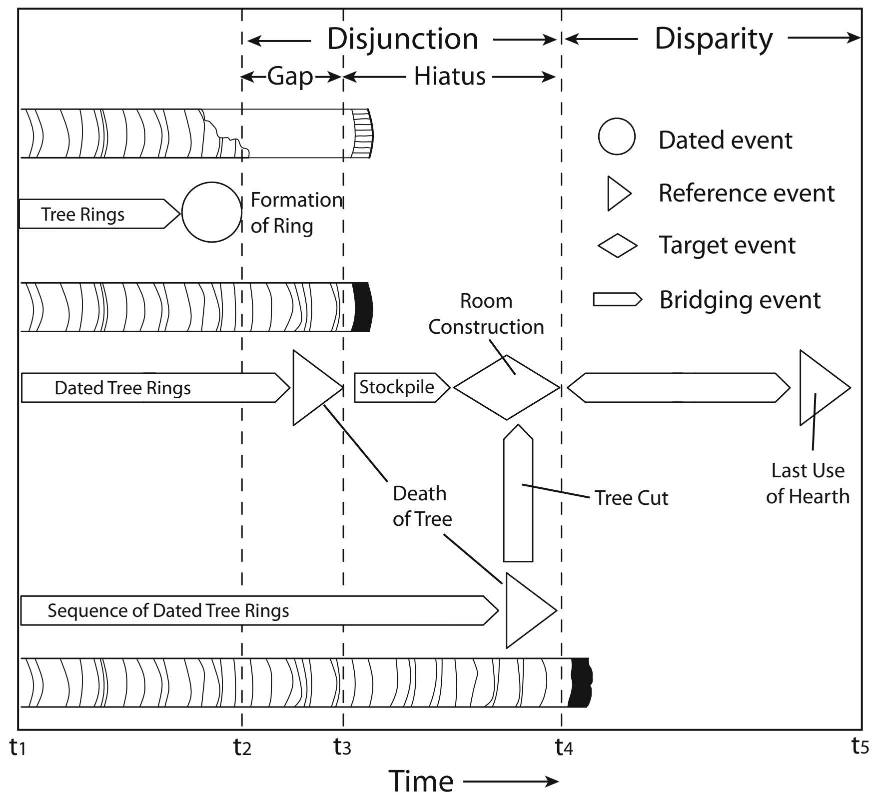


Fig. 20.2 Different kinds of dating discrepancies, using the example of dendrochronology, but applicable to all chronometric methods. Times on x-axis are t_1, initial growth of tree or limb, t_2, date of last preserved ring, t_3 and t_4, cutting dates, and t_5, date of last firing of hearth. Disjunction leads to overestimating the age of the target event, while disparity leads to underestimating its age. (Y. Salama, after Dean 1978: 227)

the opposite situation, where the reference event is later than the target event, a dating **disparity** (Fig. 20.2).

For example, let us assume that the target event is the construction of a room in a Southwestern pueblo, but the best evidence we have for its date is a reference event on the outermost preserved tree ring on a piece of wood used in the room's construction. If a dendrochronological analysis (see pp. 331–335) provides a date of AD 1321 for the dated event, *and* there is strong reason to believe that the wood was used in the original construction of the room, then clearly the construction event cannot have been earlier than 1321. However, the 1321 tree ring was not necessarily the outermost tree ring when the tree was cut, as some unknown number of tree rings could have been removed by charring, decay, or woodworking, while the wood, once cut, may also have been stored for a few years to cure before its use. In this case, the disjunction consists of a **gap** (the time between the cutting date and the date of the outermost preserved ring) and a **hiatus** (the time from the cutting date to the use of the wood in construction).

On the other hand, if we are wrong about the wood being used in the original construction, and it actually came from a later repair, then we have an instance of disparity.

Not all archaeologists subscribe to Dean's terminology, some referring, in the case of radiocarbon dating in particular, to "inner wood errors" or "inbuilt age" (IA) for cases of Dean's "gap" and "old wood offset," for his "disjunction."

20.2 Types of Dating and Their Uncertainties

Uncertainties in chronometric dates vary considerably and are not only due to biases of the kind that disjunctions and disparities characterize, but also to uncertainties in the dating methods themselves. When we have strongly confirmed historical information, for example, we can sometimes narrow down the date of an event to a single year, such as the construction of a church that is well documented. At other times, we may be able to define a specific range of date, which really involves dating two events, the beginning and end of the range. For example, it is possible to determine manufacturing ranges for much of the stoneware and china made in potteries of Staffordshire, England, by maker's marks, shipping invoices, factory records and other sources (Samford 1997). That allows us to model the date with a simple probability density function as in Fig. 20.1b, unless we have more detailed historical information on pottery production or import that allows more realistic models, with peaks and valleys in the graph to indicate production fluctuations (Fig. 20.1e).

In general, chronometric dating methods on interval scales either provide point estimates of a dated event, as in dendrochronology or varve chronology of lake sediments (Lamoureux 2002), or involve probability statements, as in radiometric dating and even most kinds of historical dating.

Both are vulnerable to the kinds of errors that Dean outlines, but the latter also entail statistical and other errors inherent in the methods themselves.

20.3 Chronology from Historical Manufacture Dates

In some kinds of archaeology, historical records and well-dated artifacts have long provided a basis for chronometric dating. The classic example is dating deposits by coin finds (see Chap. 13). Because the manufacture dates of coins are sometimes well known—often even inscribed on the coins themselves—they can be excellent for providing *termini post quem* for the deposits in which they are found. However, use of evidence from coins still requires some caution (Lockyear 2012): not only were some coins and tokens intentionally misdated (generally, to suggest they are older than they actually are), others stayed in circulation or were hoarded for long periods, so their date of deposit could follow their real or fictive manufacture date by many years (disjunction). In the case of Roman coin hoards, Lockyear (2012: 203–207) nicely shows how we can estimate the probability that a hoard was deposited in any year following the latest dated coin in the hoard. Furthermore, because they are small, they are vulnerable to displacement by bioturbation so that a single coin could be intrusive or residual (see p. 323).

Historical archaeologists have also used well-dated artifacts, such as ceramics and bottles, to date deposits. As mentioned above, some of these, such as Staffordshire table china, have fairly well documented ranges of manufacturing dates. For some other kinds of artifact, such as nineteenth-century bottles, rapid changes in bottle-making technology allow us to identify *termini post quem* on the basis of patent records for these innovations (e.g., Lorrain 1968). Traditionally, archaeologists often analyzed the chronological distributions of artifacts found at a site or in a deposit by simply plotting the ranges on a graph (Fig. 20.3). The combination of several *termini post quem* and overlaps between distributions then give a sense of the most likely date of an assemblage.

Another approach has been to calculate a "Mean Ceramic Date" (MCD). Stanley South (1977) pioneered MCD to estimate the most likely periods of occupation at historical sites in the eastern United States. It is a simple measure that involves counting all the sherds that can be identified with datable ceramic types, multiplying the counts of each type by the mid-point of the production range for that type, and then dividing by the total number of sherds of all datable types. Thus, it is an average of the mid-points of the production ranges (*not* average or median production ranges). Other archaeologists adapted this to use either mass of sherds or MNV instead of counts (although note that MNV is biased in favor of rare items, see Chap. 7). Most users have unfortunately treated MCD as a point estimate, ignoring or misinterpreting its statistical errors, which are likely to be quite large (Wesler 2014).

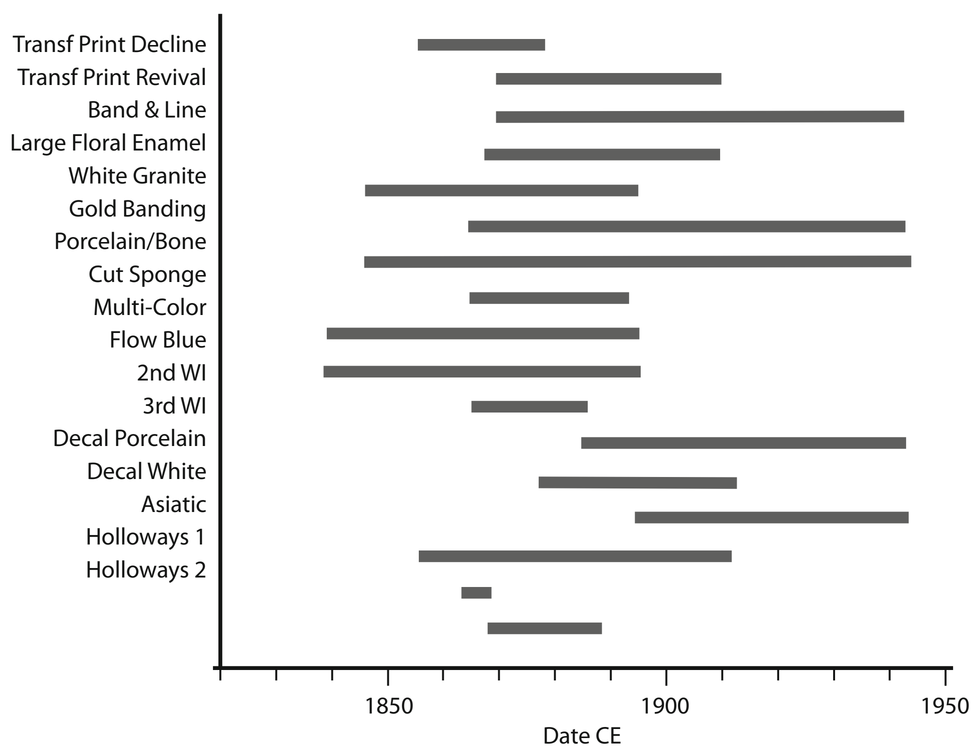


Fig. 20.3 Manufacture durations for a selection of pottery types found at the site of an historic store at Cossack, Western Australia (data from Nayton 1992). This is a simplified version omitting bottles and some of the ceramic types in the original graph. Note the greatest overlap of types in the 1880s and 90s

More importantly, MCD has several significant problems. It ignores most of the information in the production ranges by lumping all the sherds of long-lived types into a single year and performs poorly at sites that were intermittently occupied and in time ranges when the use of the pottery was rapidly increasing. It also tends to be over-influenced by abundant but rather poorly dated types, whose production ranges can often exceed a century. Worse, South tried to compensate for this last shortcoming by assigning arbitrary dates to types that had really long production ranges. Clearly, a better solution is to allow the most “diagnostic” types, those that had short manufacture ranges, to have greater influence on the result.

One way to do this is with summed distributions of the production ranges using the best information available to us (usually uniform distributions for each type, as in Fig. 20.1b). This is somewhat like “stacking” these uniform distributions, with the areas of each proportional to the number, mass, or EVE of sherds (see Chap. 7). This will usually result in peaks and valleys, with the peaks, much like the overlaps in Fig. 20.3, suggesting likely occupation periods (Fig. 20.4; Banning 1997; Ortman 2016; Steponaitis and Kintigh 1993). Alternatively, we could use Gaussian models, as in Fig. 20.1d, if there was reason to think that the production of a pottery type gradually increased to a peak and then fell off before disappearing. If we have better information, documentary evidence on production or sales of different pottery types or makers’ marks may provide more realistic probability-density distributions, as in Fig. 20.1e. Because these are probability distributions, it is possible to calculate confidence for any time interval but, unless we restrict the analysis to the short-lived types, these intervals are likely to be rather broad. These models can also be used for apportioning artifacts among periods at sites with multiple occupations (Roberts et al. 2012).

However, as with all chronometric methods, there could be disjunctions that we should take into account. For example, the pottery types that contribute most to the peak in the distribution could have seen several decades of use, reuse and curation before they were deposited. Consequently, when the artifact types used for dating may have had markedly different use-histories—for example, when clay pipes were cheap and disposable but fine china passed from one generation to the next—this is another confounding variable that would affect both MCD and the summed probability distributions. In addition, the summed probability distributions would be vulnerable to at least some of the challenges to validity that affect summed radiocarbon distributions (pp. 344–345).

20.4 Dendrochronology

Dendrochronology depends on the phenomenon in some species of trees that the thickness of annual xylem growth depends on environmental conditions. As noted in Chap. 16, this results in alternating dark and light bands, or “tree rings,” in trees growing in temperate regions. In conifers, these rings consist of an inner band of large, lighter-colored cells and an outer band of narrower, thick-walled and darker cells (p. 278). These differences are due to the fact that the tree is dormant in winter and highly active in spring, and tree rings also vary in thickness from year to year, in response to environmental changes. Trees that lack this characteristic—with rings of uniform thickness or whose thickness only decreases with age are called “complacent,” while those whose rings vary in thickness because of annual environmental differences are called “sensitive.” Under ideal conditions, sensitive tree rings form a series of thickness variations that is closely comparable to series in other trees from the same region. This allows cross-dating of trees so that we can build up a long chronology from overlapping series (Fig. 20.5).

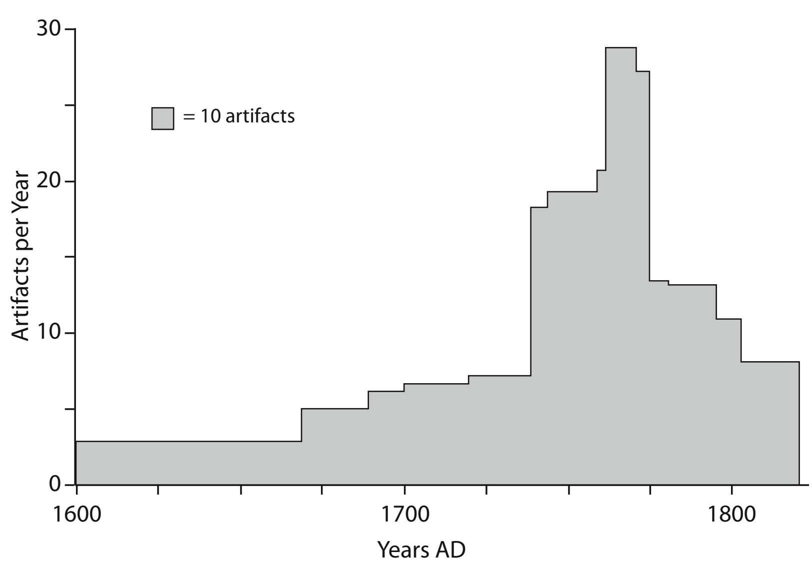


Fig. 20.4 Summed probability plot of the uniform distributions of several pottery types at the Hepburn-Reonalds Site (data from South 1977). The peak corresponds well with the historically known occupation range of 1734–1776, while an unadjusted MCD is 1744

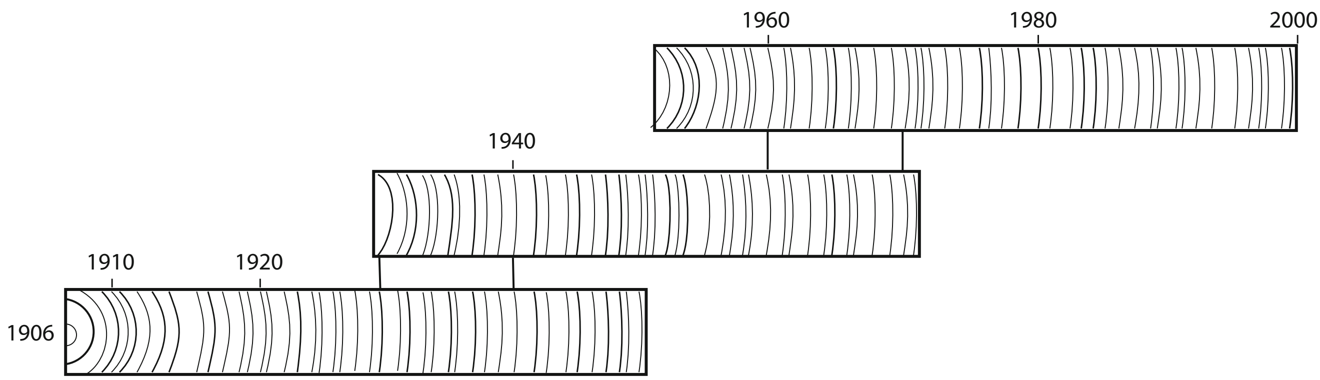


Fig. 20.5 Cross-dating tree-ring sequences by matching between separate tree specimens to obtain a longer regional sequence

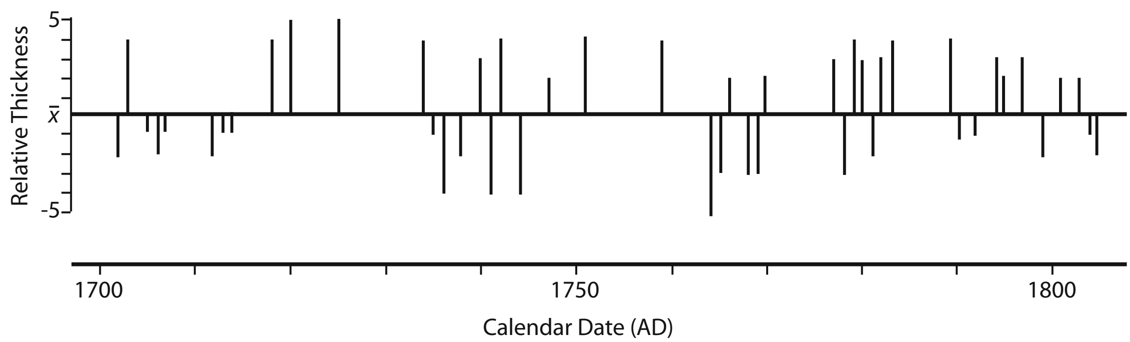


Fig. 20.6 Example of a "Skeleton plot." Instead of showing absolute thicknesses of rings, which can vary for a number of reasons, these show the deviation of ring thicknesses from a moving average (the horizontal line). Large upward bars indicate unusually thick rings, and large downward bars unusually thin ones. Where there are no bars, thickness is about average

To carry out dendrochronological work, it is necessary to prepare samples so that their transverse or radial section is very clear, and then to measure the thickness of each ring and record it on a "skeleton plot" or dendro-chronograph (Fig. 20.6). This displays the ordered growth increments as departures from the moving average of ring thickness, which standardizes them to compensate for variations in absolute thickness from tree to tree and from inner and outer parts of the same tree (xylem growth slows down as the tree ages). By comparing such plots, we can align them so that the peaks and valleys match up. Today, computer software assists in recording and displaying the variations in ring width and correlating different sequences (Litton and Zainodin 1991; Lamont-Doherty n.d.).

In regions where trees are sensitive to seasonal variation and wood preservation is good, dendrochronology provides the very best dating evidence that prehistorians could hope to have. As with any dating method, it is vulnerable to errors from disjunctions and disparities, but careful use, with multiple examples, attention to the probability of wood stockpiling, use of deadfall for fuel, or re-use of old timbers, and support from relative dating from wall-bonding or stratigraphy, can result in exceptionally fine-grained chronologies.

However, it is not immune to uncertainties, and sometimes comparing skeleton plots of several sequences or with a master sequence yields more than one possible alignment. Bayesian methods can help us evaluate the probabilities of each candidate date to determine the one that is most likely, and which ones may be nearly as likely (Buck et al. 1996: 342–352).

20.5 Radiometric and Physics-Based Dating in General

Although radiocarbon dating has some of its own peculiarities, it is just one of a suite of dating methods that depend on a physical process that provides a "clock" with which we can measure the passage of time since some event that started the clock. Besides radiocarbon dating, uranium series dating, potassium-argon and argon-argon dating, and several others employ radioactive decay as their clocks. Some others, such as thermoluminscence (TL) and optically stimulated luminescent (OSL) dating, use the gradual accumulation of energy within crystals as their clocks, the energy itself coming from various sources of radiation. Obsidian

Chronology at Pot Creek Pueblo

Pot Creek Pueblo is a site on the Taos campus of Southern Methodist University, in the northern Rio Grande region of New Mexico. It was the subject of many excavation seasons starting in 1957, mainly for the SMU field school, which collected many pieces of wood for dendrochronological analysis as well as evidence for the relative chronology of pueblo rooms on the basis of wall bonding. The latter involves the assumption that mud walls that are bonded at their junctions were built at the same time, while walls that abut them must have been added later.

Crown (1991) used the evidence of stratigraphy, wall bonding and dendrochronology to create the first detailed chronology for the construction events in the pueblo's room blocks. She made some assumptions about the stockpiling of timbers (hiatus) and conservatively restricted the use of dendrochronology to "cutting dates," that is, cases where the presence of bark indicates there is no gap. This, however, omits some evidence so that some rooms could not be dated. Beckwith's (2017) reanalysis of the sequence restores some of the dates that have gaps and makes somewhat different assumptions about stockpiling and room repairs.

Among the problems to address are cases with different tree-ring dates from the same room. The differences could be due to some combination of stockpiling or re-use of old wood (hiatus), preservation of outer rings (gap), and post-construction repairs (disparity). However, the

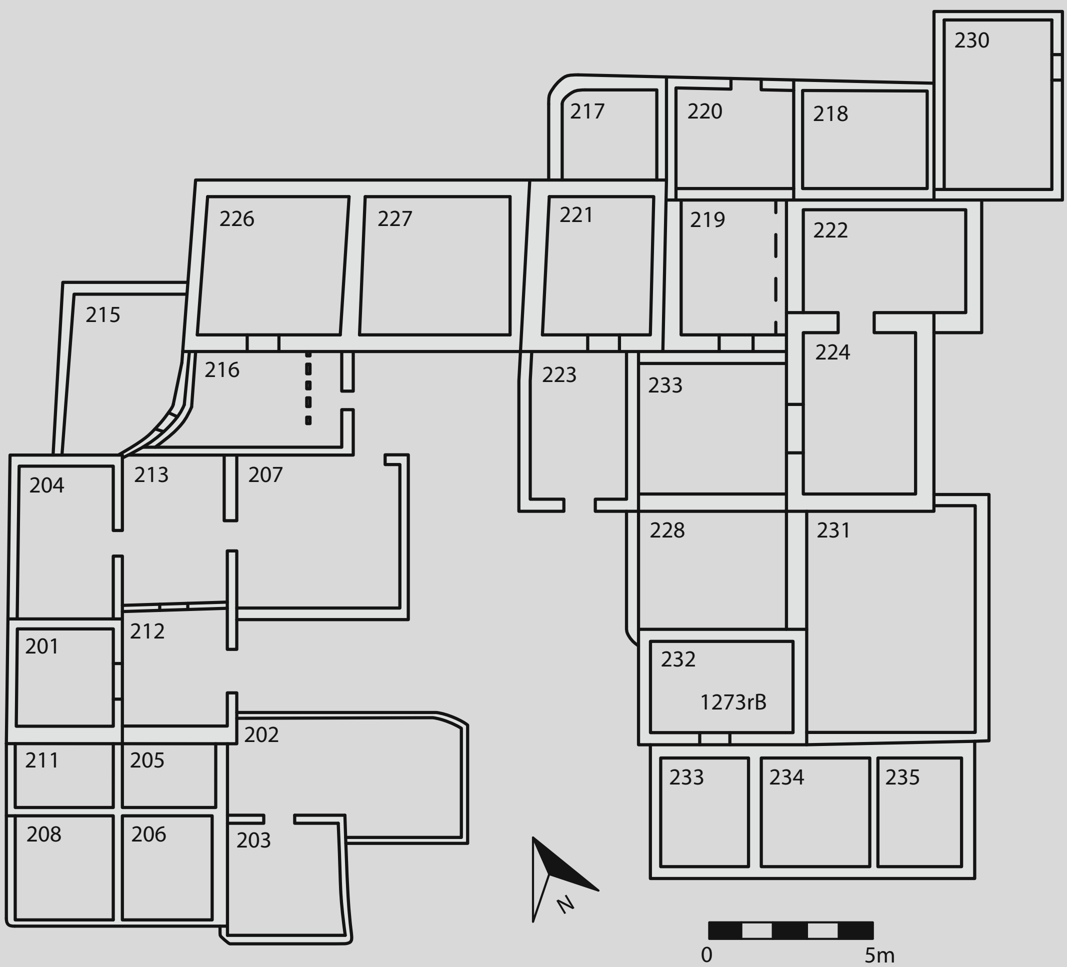


Fig. 20.7 Plan of Roomblock 2 at Pot Creek Pueblo, showing wall bonds and abutments and room numbers. (After Beckwith 2017: 154)

(continued)

Table 20.1 Two sequences of groups of cutting dates (B with bark, r with outer ring), probable cutting dates (v), and non-cutting dates (vv, +, and ++), using codes of the Laboratory for Tree Ring Research. Room numbers in each sequence are ordered from youngest at top, and probable and certain cutting dates are in bold. Dates from Crown (1991: table 1)

Room number	Tree-ring dates		Interpretation
207	**1300v** 1292vv	1298vv 1287vv	Construction after 1292 with repair after 1300
213	**1311 + r** **1273v**	1292vv 1268+ +r	Construction either in or shortly after 1273 or after 1292 with repair in or shortly after 1311
216	1292v **1285r**	1290vv	Possible construction shortly after 1285 with repair after 1292, although evidence from 213 suggests construction could have been as early as 1273
204	No dates		Construction after 1272
201	1298vv	**1272rB**	Construction shortly after 1272 with repair after 1298
226/227	**1302rB** **1273rB** **1273rB** **1273r**	**1298rB** **1273rB** **1273r** **1273r**	Tight cluster of cutting dates indicates likely construction date in or shortly after 1273 with repair in or shortly after 1302
221	1294vv **1287r** 1285 + r **1268r** 1239vv	**1288rB** 1287vv **1273v** 1262vv	Construction probably in or shortly after 1273, with repairs in or after 1288 and after 1294

distributions of these dates, when combined with the relative chronology from wall bonding, provide useful clues. A reasonably tight cluster of non-cutting dates that includes the latest date for the room is likely to be related to the construction event, but with an unknown but possibly small gap, so that these dates provide useful *termini post quem* for room construction. Cutting dates with a range of up to 5 years are likely to result from stockpiling prior to construction, while broader or bimodal distributions are likely to signal timber recycling or later repairs (Ahlstrom et al. 1991).

Turning to a portion of Roomblock 2 at Pot Creek (Fig. 20.7), for example, we can put the tree-ring dates into the framework of the relative sequence on the basis of room abutments to demonstrate how this can work. Using the ">" sign to mean "older than," we can infer the following:

201 > 204 > 215/216 > 213 > 207
221 > 226/227 > 216

If we then examine the dates from the contexts in these two sequences, we have sets of dates shown in Table 20.1.

The interpretations in the right column of Table 20.1 are of course far from certain (and not always in agreement with Crown or Beckwith), but the construction order of the rooms at least constrains the possibilities in ways that might make a Bayesian analysis useful (see pp. 341–343). When taken as a whole, 1273 CE stands out as a year of major construction activity in this part of the site.

hydration dating employs as its clock a diffusion process, the gradual thickening of a hydration layer as moisture diffuses into the volcanic glass.

What these have in common is an event that "zeroes" the clock, such as the death of an animal or the formation of a tree ring (radiocarbon), the emptying of stored energy through heating (TL) or solar bleaching (OSL), or the flaking of obsidian to expose a fresh surface (obsidian hydration).

Some clock-like processes are linear, but many of them are exponential. The radioactive decay of the isotope ^{14}C, for example, exhibits a diminishing slope with time (Fig. 20.8). Because the shape of these decay curves, at least for radioactive decay, are well understood, they make very effective clocks. We can characterize them by their mean-life (the average time that a radioactive atom will take to decay) or by the term most familiar to archaeologists: half-life. The half-life is the time it takes for half of the radioactive isotopes of an element to decay, which is also the time it takes for its radioactivity (number of disintegrations per minute per gram) to fall by one-half.

Fig. 20.8 Exponential decay of ^{14}C

20.5.1 Radiocarbon Dating

A radiocarbon date is actually not a date at all. It is a statement about the amount of radioactive carbon-14 (^{14}C) found in a sample, and consequently provides only an indirect measure of the passage of time, and then only under a number of assumptions. As radiocarbon dates are probability statements, not point estimates, radiocarbon dating has been most common in prehistoric archaeology, but it can also be useful in historical archaeology (Thompson et al. 2018).

When he pioneered this method in the 1940s, Willard Libby (1955) recognized that the radioactive decay of ^{14}C might form the basis for dating but made the following simplifying assumptions for the method to be workable.

- The amount of ^{14}C in the atmosphere in the past was the same as it is in the present, and is the same everywhere in the world
- Living organisms stop absorbing ^{14}C when they die (or tree rings stop absorbing it when the ring is fully formed)
- Living organisms absorb the isotopes of carbon without altering their ratios
- Atmospheric carbon is the only source of carbon in the samples
- ^{14}C in a sample is measurable with adequate precision and accuracy
- ^{14}C decays with a half-life of 5548 years

If all these assumptions were true, living organisms would all contain carbon with a radioactivity (beta decay) of about 15 disintegrations per minute (dpm) per gram of carbon, the activity that was prevalent in the 1940s, and a sample of carbon from an organism that died 5548 years ago would have half as much radioactivity, or about 7.5 dpm per gram. A sample that was 11,096 years old would register about 3.75 dpm per gram. This **exponential decay** would appear as in Fig. 20.8.

Conventionally, and to make modern carbon dates comparable to ones measured years ago, radiocarbon labs still report “dates” somewhat as though these assumptions were true. We call the scale on which we report these dates “radiocarbon years BP,” where “BP” or “Before Present” means “before 1950” and the “date” is really a statement about what date we would associate with the amount of ^{14}C in the sample if we were using Libby’s half-life.

However, subsequent research demonstrated the following problems with Libby’s assumptions.

- The amount of ^{14}C in the atmosphere has actually fluctuated considerably over time (secular variation), and differs slightly in the northern and southern hemispheres
- **Fractionation** occurs during organic processes, such as photosynthesis, that alter the ratios of carbon isotopes (^{12}C, ^{13}C, and ^{14}C)
- Some organisms absorb much or all of their carbon either from ocean water (which has different ^{14}C content than the atmosphere) or from inorganic sources, such as limestone

- "Background" noise can make accurate and precise measurement of the ^{14}C difficult in small and very old samples
- The half-life of ^{14}C is close to 5730 ± 30 years, not 5548.

Fortunately, scientists have found ways to account for all of these contraventions of Libby's assumptions, but they all add complications that contribute to the complexity of interpreting radiocarbon dates. One of the key improvements was to calibrate the radiocarbon chronology by making ^{14}C determinations on small groups of tree rings of known date and using large numbers of these to construct a "calibration curve," one of dendrochronology's major contributions. This allows us to check on the fluctuations of ^{14}C in the atmosphere over the past 14,000 years, while records on other materials, such as coral, allow us to calibrate even earlier dates, although with less precision.

Another important advance was to augment, and largely supplant, the older method of counting disintegrations of ^{14}C atoms when they decay, to counting the ^{12}C, ^{13}C and ^{14}C atoms themselves with Accelerator Mass Spectrometry (AMS). This advance made it possible to date smaller samples and materials of greater age (Banning and Pavlish 1978; Bayliss 2009; Bennett et al. 1977; Taylor and Bar-Yosef 2014).

20.5.2 Sources of Error in Radiocarbon Dating

Taylor (2014a) summarizes four basic sources of error in radiocarbon age determinations:

Contextual sources of error have to do with the physical relationship between the material actually dated and the event or archaeological context that an archaeologist would like to date, as in Dean's (1978) difference between dated event and target event. Most anomalous ^{14}C age determinations stem from failure to define this relationship properly (Taylor 2014a: 43). Contextual errors include those due to incorrect or poorly described stratigraphic relationships, bioturbation and other unrecognized disturbance processes (see Chaps. 17, 19).

Compositional sources of error are variations in ^{14}C concentrations or $^{14}C/^{12}C$ ratios in sample materials that result from fractionation or contamination. Contamination happens when carbon compounds that are not indigenous to the original organic material has entered the sample. Usually this is modern carbon inadvertently added to the sample during its extraction from the archaeological context or in subsequent storage, which makes the sample appear younger than it actually is, but sometimes it can be extremely ancient carbon (e.g., from oil, soot or coal) that makes it appear older. Because of the exponential nature of the radiocarbon decay curve, it does not take very much modern carbon to bias the date of a very old carbon sample (Taylor 2014b: 139). Fractionation changes isotopic ratios during certain natural processes, such as photosynthesis. Photosynthesis preferentially incorporates "lighter" isotopes (^{12}C and ^{13}C), resulting in differences in the $^{14}C/^{12}C$ ratios that have nothing to do with age. Further fractionation can result from chemical processing and measurement of samples, although dating labs routinely correct for this. Sometimes, laboratories can reduce or eliminate errors from contamination by chemical pretreatments and isolating carbon compounds that are most likely to be associated with the original organism (Brock et al. 2010).

Systemic sources of error are those stemming from physical assumptions underlying the radiocarbon method, effects of violating those assumptions, and calibration of the radiocarbon data. They include variations in the hemisphere-wide atmospheric initial ^{14}C concentrations in living organisms, differences between atmospheric and oceanic carbon reservoirs, and local offsets. Most of these variations are called "secular variation effects" and can result from differences in the production of ^{14}C over time with variation in cosmic radiation in the upper atmosphere, changes in Earth's magnetic field, and other factors. A more recent source of systemic error has been anthropogenic, as humans have released huge amounts of ancient carbon from fossil fuels since the industrial revolution (fossil fuel effect), and then created large amounts of new ^{14}C from atmospheric nuclear bomb tests during the 1950s and 1960s. The latter is usually called the atomic or bomb effect. Reservoir effects are due to the fact that some organisms absorb carbon from sources other than the atmosphere, such as the sea or limestone.

Measurement sources of error depend on the ways in which a radiocarbon lab has determined the ^{14}C concentrations in the samples. They include whether the lab estimates these concentrations by beta decay or Accelerator Mass Spectrometry (AMS), how they have accounted for background noise, whether the instrument was operating properly, potential laboratory contamination, mathematical errors, and potentially even mis-labelling of specimens. When a radiocarbon lab reports a result, it should include information that allows users to assess these measurement issues (see "Reporting Radiocarbon Results").

20.5.2.1 Calibration of Radiocarbon Results

Calibration is the process that allows us to correct for many of the errors just mentioned, especially the systemic ones. It is based mainly on the fortunate availability of good tree-ring sequences built up from living trees and ancient wood in several parts of the world that allow us to make radiocarbon determinations on tree rings, individually or in small groups, of known date. Radiocarbon scientists have also been pushing the calibration to much earlier dates by augmenting the dendrochronological records with sequences from varves (thin annual layers of sediment in glacial lakes), corals and ice cores (Geyh and Schlüchter 1998; Jöris and Weninger 1998; Kitagawa and van der Plicht 1998).

After several decades of research on radiocarbon calibration, we have excellent calibration curves—graphs that show the relationship between uncalibrated "radiocarbon years" and calendar years—for atmospheric carbon in both northern and southern hemispheres, as well as curves for marine carbon. At time of writing, the most recent curves available are IntCal13, resulting from the 2013 meeting of the IntCal Working Group (Reimer et al. 2013), although we anticipate the IntCal20 curves very soon (Jones 2020).

A calibration curve is a graph of the radiocarbon "age" against calendar years. If the two were perfectly correlated, differing only because of Libby's mistaken meanlife and failure to account for fractionation, we would see a straight line (Fig. 20.9), and the normal or Gaussian distribution of statistical error about the radiocarbon determination would result in a similarly Gaussian probability distribution of the calendar date, but centered on a different mean. However, the calibration curve is not linear, but has "plateaus," steep areas and "wiggles." Consequently, the calibrated date only comes close to having a symmetrical, Gaussian distribution if the uncalibrated determination has a small error and happens to intersect a smooth portion of the calibration curve (Fig. 20.10). If that portion is "steep" enough, it might also seem to have a smaller error than the original determination (although they are in different units). If, instead, the determination intersects a plateau or wiggles in the curve, there is more than one solution, and the resulting probability density function is multimodal and rather broad, sometimes leading to a wide range of possible calendar dates (Fig. 20.11).

While no one should try to calibrate radiocarbon dates just by looking at the calibration curve, this demonstration illustrates the value in checking the radiocarbon curve before you submit your samples to see if there are any plateau areas that would likely affect your results. Many radiocarbon labs provide the option of "high-precision" dates, at a higher cost, but there is no sense in incurring such cost if the result is likely to occur in one of these plateau areas.

Today, we calibrate radiocarbon determinations with software. Most of the available software not only carries out the calibration, providing us with calendar-year date estimates within a specified credible interval (e.g., 95% or 68%, see pp. 139, 343), but also a graph of the probability density function and a set of tools to help us to understand the relationships among sets of dates or to test statistical hypotheses about the dates. Prominent among this software are OxCal, BCal, Datelab, ChronoModel, and Calib (Bronk Ramsey 1998, 2018; Buck et al. 1999; Jones and Nicholls 2002; Lanos et al. 2016; Stuiver et al. 2019).

20.5.2.2 Reporting Radiocarbon Results

By international convention, and to ensure quality, publication of any radiocarbon result should *always* include the laboratory identification number and several other things. These include the "percent modern" (the percentage of ^{14}C remaining, relative to the amount expected in the atmosphere in 1950) or the "determination" or uncalibrated "date" BP (before present, i.e. before 1950), the statistical error (one standard deviation, typically represented as "σ" or "sigma"

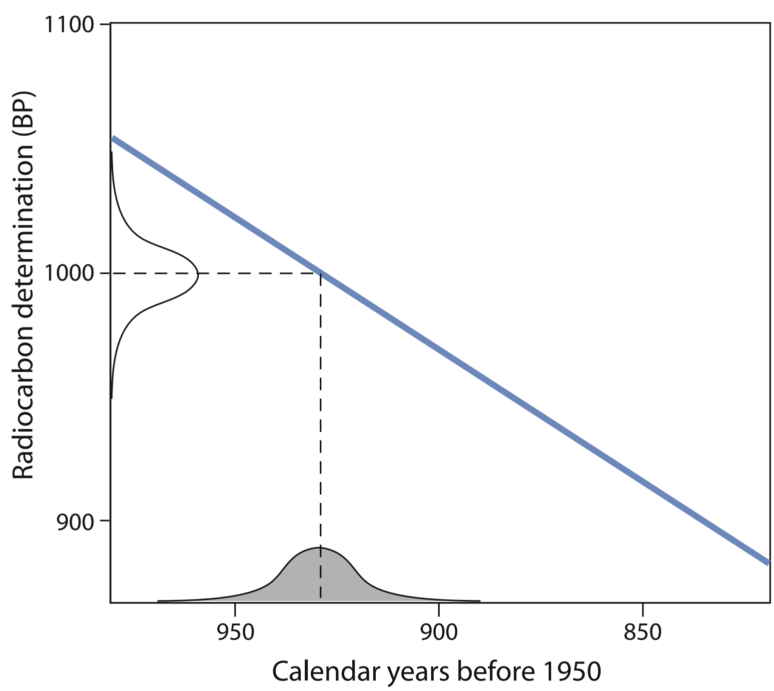


Fig. 20.9 Unrealistic, linear calibration relationship between radiocarbon "years" and calendar years. Both probability distributions are normal, but the breadth of the calibrated one depends on the slope of the line

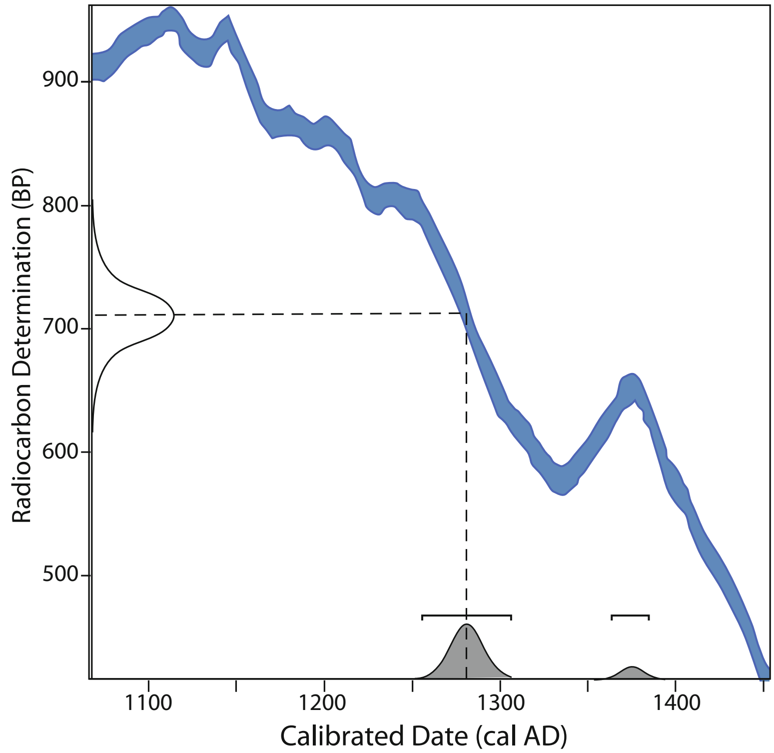


Fig. 20.10 A more realistic radiocarbon calibration for a determination of 710 ± 30 BP. Because the radiocarbon date intersects a fairly steep and smooth portion of the calibration curve, the calibrated date only slightly departs from a normal distribution, except for a small secondary peak around AD 1375 that results from a bend in the calibration curve. (Modified output from OxCal v4.3.2, Bronk Ramsey 2018)

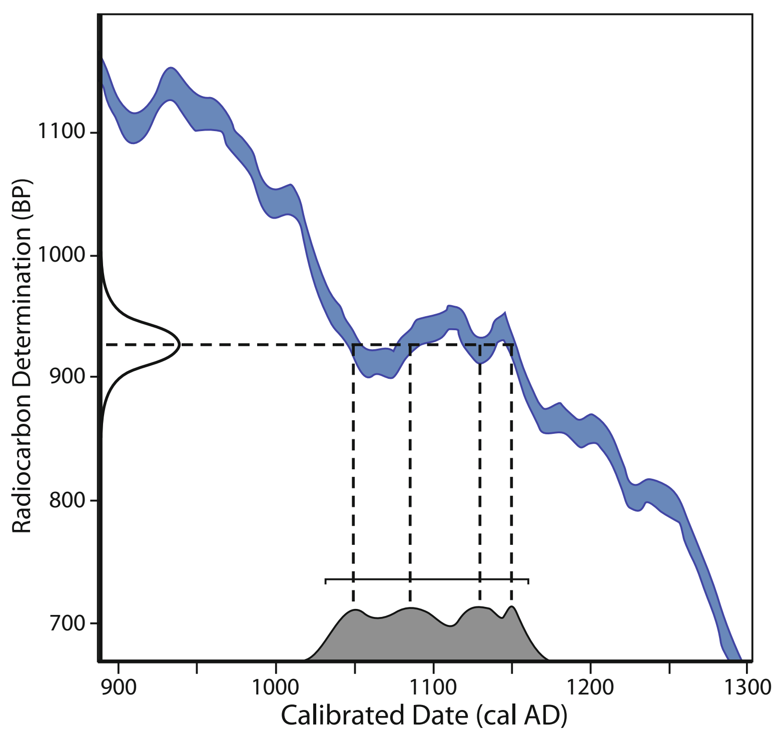


Fig. 20.11 Another radiocarbon date with a smaller statistical error than in Fig. 18.10 (927 ± 25 BP), but intersecting a "plateau" with "wiggles" in the calibration curve. The resulting calibrated date distribution is broad and multimodal. (Modified output from OxCal v4.3.2, Bronk Ramsey 2018)

even though it is a sample statistic, not a population parameter), and the measured or estimated fractionation (δ^{13}C or "delta-^{13}C"). For relevant samples, it should also include any reservoir correction ("delta-R") that was made (Bayliss 2015: 682–687) and, for bone collagen, the collagen yield. These data are unaffected by any revisions to the calibration curve. For calibrated dates, always tell readers what version of the calibration curve you used (e.g., IntCal13 atmospheric), what the credible interval is (e.g., 95% or 68%), and express the date as "cal BC," "cal AD," "cal CE," or "cal BP". The last is commonly found in geological publications, while most archaeologists favor the other versions. Make sure that you identify the material that was dated, such as oak charcoal, deer bone collagen or squash seeds, to allow readers to distinguish short-lived samples from ones that could have large disjunctions. It is also good practice to report what portion of the material was isolated for dating (e.g., bone collagen, total carbon, humic acid), and what chemical pretreatments were used to clean the samples and extract carbon.

20.5.2.3 Wiggle-Matching and "Bomb Effect"

Problems that result from the vagaries of the calibration curve—plateaus, wiggles, and multiple intersections—are especially prominent in the portions of the curve following the industrial revolution and the beginning of atmospheric atomic bomb testing. The former, through the burning of enormous quantities of coal, dumped huge quantities of very old carbon, containing no ^{14}C at all, into the atmosphere. The latter, by irradiating nitrogen atoms in the atmosphere, created large amounts of ^{14}C to supplement the "normal" amount that results from cosmic-ray collisions with nitrogen. These at first glance would appear to make radiocarbon dating effectively useless for samples later than about AD 1780, and this has been a disincentive among historical archaeologists. However, creative researchers realized that there were ways to take advantage of unusual fluctuations in the calibration curve.

Until recently, most radiocarbon determinations on wood or charcoal, including those used to make the calibration curve itself, employed samples that included five or ten successive tree rings, which averages out the annual variations in radiocarbon abundance in the atmosphere. Dating the individual rings separately can "fingerprint" different parts of the curve by the "wiggles," thus allowing us to distinguish between dates that would otherwise look identical once calibrated. This "wiggle-matching" is costly but is the only way to resolve chronology in those tricky parts of the curve (Bronk Ramsey et al. 2001).

As noted, most radiocarbon determinations are relative to the year 1950. One reason for using this date was that atomic testing radically changed atmospheric radiocarbon abundance after that date. However, somewhat as with wiggle-matching, it turns out that we can use that "bomb effect" to date materials of the last 70 years with great precision, now that we have records of recent levels of atmospheric radiocarbon, which peaked around 1965 and then started to decline (Hua et al. 2013). Among the many non-archaeological applications of this effect is the detection of counterfeit rare whiskies (Dunbar et al. 2018).

20.5.2.4 Sample Type and "Chronometric Hygiene"

With respect to the kinds and severity of errors mentioned above, not all potentially datable materials are equal. This has led to some researchers "weeding out" existing radiocarbon determinations as inappropriate or potentially biased sources of chronological information (Fitzpatrick 2006; Spriggs 1989). Different potentially datable materials have different potential advantages and flaws. Wood and wood charcoal are materials commonly used in radiocarbon dating, but the fact that wood can come from extremely long-lived trees, recycled timber, or inner portions of a tree that "locked in" their ^{14}C long before the tree was felled or its branches removed has led to the "old wood" problem: a date on wood or wood charcoal can long precede the date of felling or branch removal, a problem made worse by woodworking, stockpiling, and recycling of timbers, and the possibility that charcoal has absorbed humic and fulvic acids from the surrounding soil (Dean 1978; Taylor 2014b: 69). "Short-lived" plant materials, such as seeds, papyrus, or basketry, can have advantages over wood in that their material was probably harvested in 1 year, and may be closer in age to the target event than wood would likely be (smaller disjunction). However, their $\delta^{13}C$ values may be quite different from those in wood. Marine sources (shell, bones of fish, marine mammals), on the other hand, have reservoir effects due to the fact that the concentration of ^{14}C in the sea is quite different from that in the atmosphere (Cook et al. 2015). Terrestrial snails also show reservoir effects from their uptake of carbon from carbonates in their environments that is effectively devoid of ^{14}C.

Chronometric hygiene concerns the evaluation, and often rejection, of radiocarbon determinations on the basis of their failure to meet specified standards of validity, reliability and confidence. Frequently, their rejection is due to their ambiguous or inappropriate context, but many are also rejected on the basis of their material (compositional errors).

Spriggs (1989) outlined four criteria that he thought would ensure the quality of radiocarbon evidence, describing the process of rejecting unsuitable samples as "chronometric hygiene": the sample must not be from long-lived species, multiple determinations are needed from each context, multiple determinations are needed from stratified sites, and association between the sample and the cultural context must not be ambiguous. This is a very strict protocol that unfortunately would exclude many sites and projects from being dated by radiocarbon.

Table 20.2 outlines the most common advantages of and objections to different sample materials and contexts. Very strict adherence to chronometric hygiene involves rejecting

Table 20.2 Summary of the seriousness of compositional, contextual and systemic effects on various classes of potentially datable material for the radiocarbon method. For more detailed discussion, see Taylor (2014c: 130–154)

Material	Compositional effects	Contextual effects	Systemic effects	Comments
Wood charcoal	Low	High	Low	"Old wood" problem
Charred fruit or seeds	Low	Medium	Low	Seeds are "short-lived" but can be residual in deposits
Terrestrial bone	Medium	Low	Low	Death may be butchery event Dietary reservoir effects, potential for poor collagen preservation
Marine animals	Medium	Low	High	Marine reservoir effects
Terrestrial gastropods	Medium	Medium	High	Gastropods that obtain some of their carbon from limestone have large offsets
Freshwater shell	Medium	Low	High	Reservoir effects
Marine shell	Medium	Low	High	Reservoir effects
Soil	High	High	High	Bulk-sampling effects

most wood charcoal because it comes from long-lived species and is therefore vulnerable to the "old wood" bias (disjunction). However, Fitzpatrick (2006: 393) and most archaeologists have found that close adherence to Spriggs's criteria is often untenable. Sometimes this is because wood charcoal is the only material available, or because the material is not reported, and often because there are too few dates for us to reject cases lacking multiple dates from the same context. In addition, how much risk of error is acceptable in any particular context depends on the existing quality of chronological knowledge (Taylor 2014c: 130) and the research questions asked. Strict chronometric hygiene may be appropriate where existing knowledge calls for sub-century refinements to a well-established chronology, but more forgiving protocols may serve in cases where the chronology is poorly or only moderately understood.

Bulk-sample Effect: Before AMS dating was available, archaeologists sometimes combined bits of charcoal from the same context to provide a pooled or bulk sample. This may have been understandable then but should be avoided today. There is a non-negligible probability that any seed, bone or piece of charcoal in a deposit might be residual, and we can detect those residuals as outliers when we get individual dates on several items that we can subject to statistical analysis. However, if we pool them and only get one date, there is no way to make that assessment and the probability of obtaining a biased date increases with the number of items we pool together (Hamilton and Krus 2017: 688).

Boaretto (2008) also suggests some basic steps to protect the validity of a radiocarbon dating program, including the following.

- Check the calibration curve for the range where you expect your dates to fall, to see if it is feasible to answer your archaeological questions
- Exclude any samples of uncertain context or unclear relationship to the research questions. Including problematic samples will only lead to embarrassing explanations later on
- Define the expected quality of every sample you will use
- Consider the sample materials. It is usually better to use short-lived samples, such as seeds, but preferably when they are in clusters in hearths or pits, rather than single seeds that might be residual or intrusive. Bone is also a good choice at sites where collagen is well preserved because the death of the animal makes a good dated event
- Characterize the sample material. For example, infrared analysis of bone collagen allows detection of any contamination by clay or non-indigenous organic materials. Raman spectroscopy of charcoal will similarly help identify contamination by humic material
- The radiocarbon lab should use extraction and purification pretreatments to remove humic materials, carbonates, and other carbon materials that do not belong to the original charcoal or collagen

Many archaeologists have additionally excluded determinations on the basis of the size of their statistical errors, excluding dates with errors greater than, for example, 200 years. This is not, in fact, good practice, as long as the determination and the sample on which it is based are otherwise appropriate (Hamilton and Krus 2017). The software for Bayesian analysis of dates already takes imprecision in errors into account. What is more important than the size of the errors is the connection between the radiocarbon sample and the target event.

Bronk Ramsey (2009) emphasizes the importance of detecting and managing outliers. Outliers can result from the sources of error already mentioned, but it is not always obvious when an error has occurred, or which determinations are in error. If we can identify the outliers correctly, we can omit them from analysis. OxCal's "agreement index" is one tool for rejecting anomalous dates. However, we can expect about one in 20 determinations to have an agreement index less than 60%, and should only omit them if there are grounds, such as suspected disjunction or contamination. Alternatively, we can take the view that we can never be certain which dates are incorrect, and instead weight the dates so that the ones most likely to be outliers contribute less to the result of the analysis (Christen 1994). This requires "flagging" some dates as potential outliers and depends on a model for their expected distribution, such as the normal distribution when the source of the outlier is probably measurement uncertainty. However, other models are more appropriate when some other source of error, such as contamination or reservoir effect, is likely. Software for Bayesian analysis of radiocarbon databases, such as OxCal or BCal, provides these tools.

20.5.2.5 Too Few Dates? Or Too Many?

One of the major questions of radiocarbon analysis is how many determinations to get. Paradoxically, in a sense it is possible to have too many, although it is more common to have too few. One of the effects of statistical processes is that, the larger the sample size, the larger the number of observations that fall in the "tails" of the distribution. Consequently, if you date an event with many samples, and then use statistical methods to sum the results (see below), your credible interval for that date will usually be broader than if you date it with only a few samples, unless the date is constrained by a model in a Bayesian framework.

As Bayliss and Orton (1994) point out, the key is to determine how many dates you need to answer a particular question at a particular level of statistical confidence. They consider a number of types of question: questions concerning a single event or phase, ones concerning the relationship between two events or phases, and ones concerning the relationships among three or more events or phases. They then formalize these questions in terms of null hypotheses; for example, the null hypothesis might be that there is no age difference between two events, while the alternative hypothesis might be that event A is older than event B (A > B).

Next, they use procedures for estimating sample size that have been discussed in Chap. 6 (pp. 100–101). Not surprisingly, a key is to tie decisions about sample size to research questions.

20.6 Bayesian Analysis of Chronology

Software that both calibrates radiocarbon dates and facilitates Bayesian analysis of date distributions, whether from radiocarbon or other methods, under a variety of prior assumptions has been revolutionizing archaeological chronology. However, as with any relatively new method, it requires thoughtful use rather than "plug-and-play" analyses (Buck and Meson 2015; Dye and Buck 2015; Hamilton and Krus 2017).

As with most research designs (see Chap. 6), at the heart of Bayesian chronological analyses are the definition and modelling of a chronological question. When was this hearth used? What was the duration of this phase of occupation? Was this site occupied before or after a Rapid Climatic Event (RCE)? Generally, it is better at the beginning of a dating project to have only a few very focused questions than to start by building a very complex model to solve many questions at once.

The next step is to select radiocarbon determinations or other dating evidence relevant to your question, make sure you understand their relationships to the target events, and build one or more models that describe the prior information as you understand it (Fig. 20.12). For example, you might have stratigraphic information that tells you that the last uses of a hearth must have occurred after the deposit of some charcoal in a layer that underlies the hearth, but earlier than the construction of a storage pit that contained a deposit of charred barley (Fig. 20.13). This has an obvious connection to Harris matrices (Chap. 19).

With the model specified formally in the software, and the dating information listed, the software builds up probability distributions (probability density functions) for the dated events by using a "Gibbs sampler" (a Markov chain Monte Carlo algorithm) to make thousands of computer simulations that take the calibration curve and your model into account (Fig. 20.14). You should repeat this process. Once you have completed at least a few runs (starting each with a different "random number seed"), check on consistency of the results. Each run will yield a result that is at least slightly different, but major discrepancies or the software's failure to provide a result usually means that your model has problems, either being too complicated or just incorrect. Whether your model

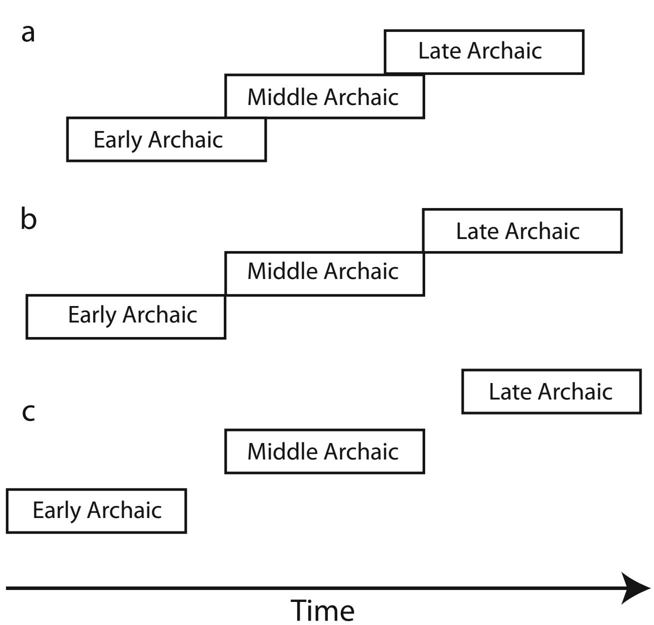


Fig. 20.12 Three potential models for the sequence from Early to Late Archaic with periods overlapping (**a**), abutting (**b**), and separated by gaps (**c**). (After Buck et al. 1996: 218)

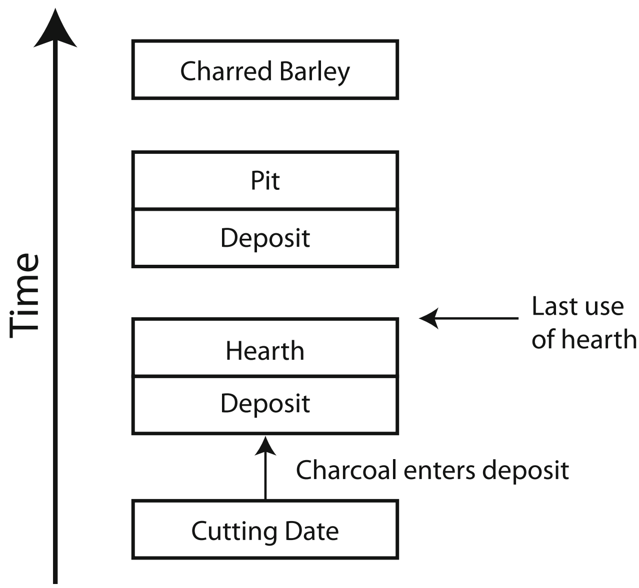


Fig. 20.13 A model for a specific dating situation, with the use of a hearth "sandwiched" stratigraphically between the deposit on which the hearth was placed and the deposit that covered it after some unknown interval. The covering deposit is cut by a pit that contained charred barley, and the underlying deposit contains some charcoal with preserved bark (making it possible to estimate a cutting date). Gaps indicate possible time lags between events, while adjacent boxes indicate that one event (such as construction of hearth) immediately followed another

is successful or not, you can later begin a new round by adding further dates and perhaps new prior information or a competing model, as Bayesian results are always work-in-progress.

Bayesian analysis entails a degree of subjectivity, in that we construct models of the data and formalize our prior beliefs about the data on the basis of archaeological knowledge. However, this does not mean that Bayesians just pick and choose dates as they please; on the contrary, Bayesian analysis provides a rigorous methodology for evaluating outliers and ensuring that results are robust. In the case of chronology, most Bayesian applications have involved constraining the chronometric evidence by prior beliefs about the stratigraphic sequence, although that is only one possible source of prior information, which could alternatively include calendar dates on, for example, coins or well-dated floods or earthquake events. In the absence of evidence for the order of the dates, it is appropriate to use a uniform prior distribution, sometimes called an "uninformative" prior. There is no justification for ordering the dates by the apparent order of their means (Buck and Meson 2015: 571). Nor should we treat our preconceptions about the dates as evidence to constrain the analysis, as that would be tautological (Hamilton and Krus 2017: 192).

Because the results of a Bayesian analysis depend on prior information, it is important to conduct sensitivity analyses to see how much the results depend on specific assumptions. Re-running the dates with revised models that selectively

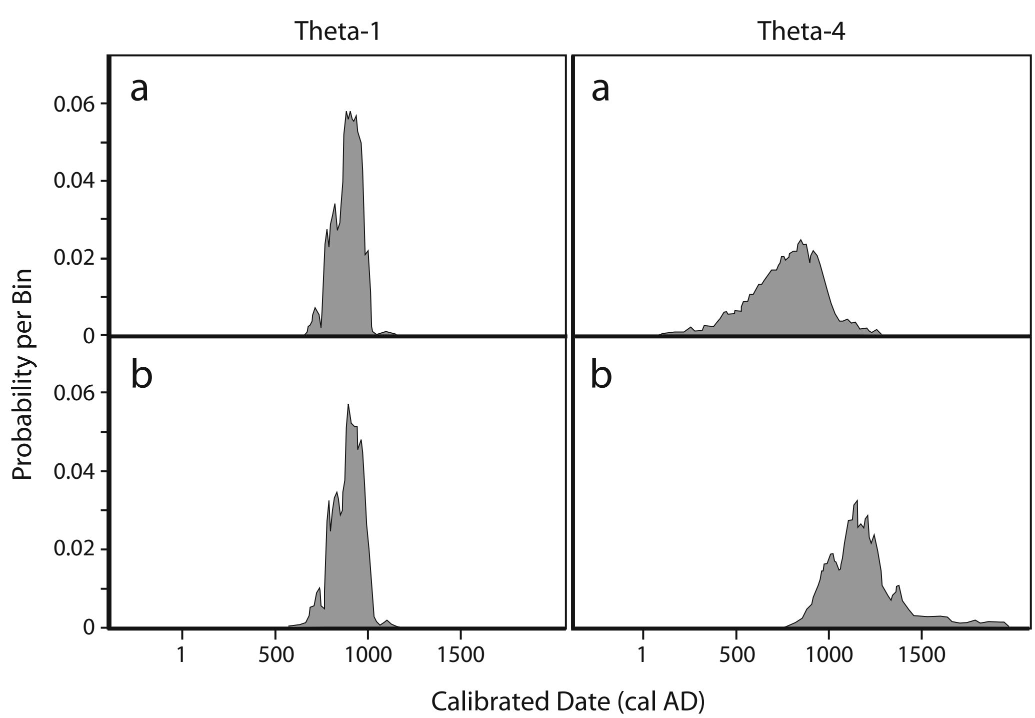


Fig. 20.14 An example of output from a Bayesian analysis of two dates called Theta-1 and Theta-4, (**a**) with no model, and (**b**) with a Bayesian model that takes stratigraphic order into account. Note that this has a particular impact on Theta-4 in this case. (After Dye 2011: 135)

remove some of the constraints or change the assumptions helps us make this evaluation (Bayliss et al. 2011).

Although most of the chronometric applications of Bayesian analysis have been for radiocarbon chronology, Bayesian modelling is equally applicable to other dating methods (e.g., Millard 2006).

20.6.1 Credible Intervals

Archaeologists have tended to have a love affair with 95% confidence intervals. Selection of confidence intervals, as with any statistical interpretation, should be based on the problem you are trying to resolve, and 95% confidence may be too conservative for some of these. You need to balance the risks of making type-I and type-II errors (p. 130).

The Bayesian analogue to confidence interval is the credible interval, which is not exactly the same thing. While, in "frequentist" statistics, 95% of the confidence intervals of repeated samples will include the "true value" of a population parameter, the credible interval treats the "true value" as a random variable and constitutes the Highest Posterior Density Interval (HPDI). This includes the portions of the posterior distribution (as in Figs. 20.11 and 20.14) that are more credible than portions outside the HPDI. That is, a 95% credible interval includes the narrowest intervals that jointly have a 95% probability of containing the parameter of interest.

One practice that you should particularly avoid is to attempt to date the duration of archaeological periods or phases as the interval between the lower end of the 95% credible interval for the oldest date, and the upper end of the 95% credible interval for the youngest date from the phase. This exaggerates the length of the period and increases the overlap between phases unrealistically (bias). It is much preferable to use the interpretive tools in Bayesian calibration software to assess the probable dates of transitions and durations of periods or phases.

20.7 Summed Probability Distributions (SPD)

Now that radiocarbon dates are sometimes abundant, an increasingly popular practice among archaeologists is the summation of calibrated radiocarbon probability distributions (SPD), much like the alternative to Mean Ceramic Date (p. 330). Tools in OxCal software (Bronk Ramsey 2009, 2017) make these summations relatively easy to generate. In most applications, as with MCD, this is assumed to provide a good "anthropogenic signal" of the period of a site's occupation, or is used as a proxy for human population or energy consumption to facilitate study of demographic change over large regions and long timescales (Bamforth and Grund 2012; Bevan et al. 2017; Freeman et al. 2018; Robinson et al. 2019; Timpson et al. 2014), or to make inferences about the spread of some cultural phenomenon in space (e.g., Crema et al. 2017).

For example, one project took 13,658 radiocarbon dates from 12 regions in western Europe to explore the relationship between human demography and both climatic changes and the introduction of farming, on the assumption that "there is a relationship between the number of dated archaeological sites falling within a given time interval in a given region (or their summed date probabilities) and population density" (Shennan et al. 2013: 3). Figure 20.15 shows the result for the entire dataset as well as for a "null model" that represents

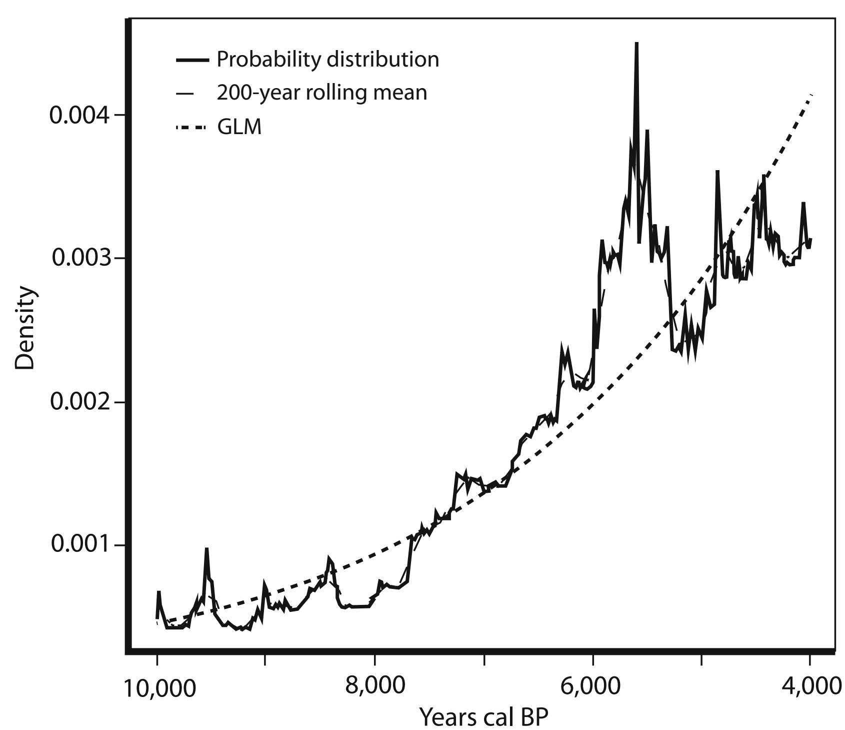


Fig. 20.15 Summed probability distributions (SPD) for a database of 13,658 radiocarbon dates in western Europe (after Shennan et al. 2013). The number of bins is 6497. The exponential General Linear Model (GLM) is assumed to model both "taphonomic loss" and general population increase over the period and serves as a basis for comparison with the SPD

the assumed effects of both taphonomic effects (older sites being underrepresented) and long-term expected population growth. As the introduction of agriculture was not synchronous over western Europe, they also produced SPD for each of the 12 regions. Notably, ten of these show significant increases in population, above the curve for the null model, about the time that agriculture appeared, followed by a substantial drop.

From the perspective followed in this book, this is an example of indirect measurement, and brings up issues of validity. Such uses of radiocarbon data depend on several assumptions, including:

- The events that radiocarbon dates reflect are associated with human activities on the site or sites (or at least those events that involved using wood, burying deceased people, or killing animals)
- The available dates constitute a representative sample of such events

and, for studies of demographic trends,

- There is a (linear) relationship between the number of such events and the number of people living at the site or sites

Many archaeologists would consider some or all of these assumptions to have at least face validity, although others might ask, "would we consider the number of stone tools or the number of house pits in an available sample to be proportional to human population sizes?" On the other hand, archaeologists often seem comfortable with using relative abundances of pottery types in this way, as with Mean Ceramic Date and its substitutes. In the case of radiocarbon evidence, specifically, some confounding factors could call some of these assumptions into question (Attenbrow and Hiscock 2015; Bishop 2015; Williams 2012: 579):

- The shapes of the SPD are influenced by the shape of the calibration curve. Steep portions of the curve tend to create spurious peaks in summed probabilities, while plateaus smooth out real peaks, making them less obvious. The poorer resolution of earlier portions of the calibration curve is also an issue.
- Taphonomic processes affect the representativeness of the sample. Older dating materials may be less likely to be preserved than younger ones ("taphonomic loss") or be buried more deeply and less accessible to archaeological investigation.
- Changes in the kinds of activities, such as food processing, or types of fuel used (e.g., from wood to dung) over the course of site occupations would complicate the relationship of available radiocarbon samples to numbers of site inhabitants
- Archaeologists tend to select materials for dating to solve specific chronological problems, as recommended elsewhere in this chapter. This selectivity must also affect sample representativeness.
- The number of items submitted to dating labs varies with research agendas, available funding, and intensity of research, potentially calling into question the representativeness of the sample available to us. In some cases, multiple carbon dates may be associated with a single event, while other events have no dates at all. Strong interest in such "hot" topics as the origins of agriculture surely also affect the relative abundance of dating evidence.
- As noted above, the dates that provide the dataset can be subject to highly varying disjunctions and disparities, so that the events contributing to the summed distributions have variable relationship to the human events they purport to represent.
- As SPD involve statistical processes, we can expect some apparent anomalies (peaks or valleys in the plots) to be simply statistical outliers or "false positives" (Timpson et al. 2014: 550–551).
- The selection of "bin" size—the interval within which the analysis groups dates—can affect the shape of the distribution (cf. histograms, Fig. 5.17).
- Sample size also affects the shapes of summed distributions.

Consequently, to have reasonable validity, archaeologists who use summed probability distributions must account for at least some of these confounding variables if they are to make convincing arguments.

Accounting for the calibration curve is certainly possible. One way to do this is to create simulated radiocarbon dates for series of calendar dates either at regular intervals, such as every 50 years, or randomly over the period of interest, or drawn randomly from a null model that takes factors like taphonomy into account, and possibly to repeat this at least 1000 times. Calibrating these simulated date-sets and making SPD from them provides a standard for comparison (with a 95% credible region in the case of multiple simulations), allowing us to see if the peaks and valleys in the SPD of real dates depart significantly from the shape of simulated SPDs. Use of a moving average of several centuries to produce trend lines also helps to remove the calibration-curve effect (Williams 2012: 584).

Attempts to correct for "taphonomic loss" have focused on the possibility that the sample of radiocarbon determinations underrepresents older events. Surovell and Brantingham (2007; Surovell et al. 2009) suggest compensating for this main effect by modelling exponential loss of sites and carbon materials over time. Imposing a simple model on the data poses some risk that it could distort or erase some genuine trends (Williams 2012: 586), so it is useful to take geoarchaeological evidence, where possible, into account to assess whether it is appropriate to assume a single loss rate over time. For example, date series on geological deposits in the same region could help us assess whether rates of losses through erosion were higher in some periods (Surovell and Brantingham 2007: 1874). As just noted, taphonomic loss can be incorporated as a null model in accounting for the calibration curve.

Some archaeologists believe a large enough sample size will compensate for the individual problems with specific dates and sites, such as changes in fuel sources or archaeologists' selection of material for dating, creating a quasi-random sample of events (e.g., Smith 2016: 214). However, it is not clear how large a sample this requires or whether typical radiocarbon databases meet the criteria for Bayesian exchangeability. Michczynska and Pazdur (2004) used Monte Carlo resampling to determine that for a 14,000-year sequence and mean error on individual dates of ±115 years, the minimum sample size that would keep statistical fluctuations below 50% was 200, while Timpson et al. (2014: 552) claim "broad scale" similarity of SPD even in very small datasets. Most reasonably successful applications of the method have sample sizes greater than 1000 (Williams 2012: 579).

In case sheer sample size is not enough to compensate for it, some researchers specifically address bias from variation in research intensity. One way to deal with this, or at least the aspect of it that some sites, regions, or periods of time may be more intensively sampled than others, is to resample the radiocarbon database in a way that reduces the impact of very large and intensive projects, typically by randomly selecting no more than, say, five dates from any one period at any one site, thus making the sample sizes from different sites and projects more comparable. Another approach is to forgo using the individual dates and instead using mean calibrated dates for the phases at sites, so that the SPD are effectively counting numbers of sites dated to each period, rather than numbers of dates.

To deal with false positives, one approach has been to assume that genuine positives should be patterned because of some underlying population process, and to remove isolated points and small groups of positives up to a maximum of 5% of the area of the SPD (Timpson et al. 2014: 551).

A sensitivity analysis, involving multiple SPD that employ different "bin" sizes, allows us to evaluate whether the choice of bin size is having undue influence on the shape of the distributions. Generally, larger bin sizes (> 100 years) tend to blur the stochastic variations in the data. Use of a moving average over 200-year intervals also accomplishes this.

While these approaches may correct for some of confounding variables, Contreras and Meadows (2014) found, in a series of simulations of summed probability distributions, that "even under ideal conditions, it is difficult to distinguish between real and spurious population patterns, or to accurately date sharp fluctuations." Timpson et al. (2014) attribute these more pessimistic results to edge effects and use of an inappropriately short time series.

Going back to a question near the beginning of this section, we might test whether an SPD has better than face validity by comparing results against other indirect measures, such as ones based on total floor area of settlements over the periods of interest, changes in the proportion of juveniles in human mortality profiles, anthropogenic impacts on pollen records, or evidence for site abandonments or population pressure (Shennan et al. 2013: 3; Surovell and Brantingham 2007: 1875–1876). It would also be useful to test for possible effects of research intensity by dividing datasets into groups by the decade in which the samples were excavated or dated.

20.8 Summary

- For any chronometric dating program, it is essential to identify explicitly the target events you want to date and relevant reference events that might be datable
- Dating discrepancies can cause bias in date estimates unless steps are taken to compensate for them, while other uncertainties in chronometric dates can affect precision, accuracy, or both
- Dendrochronology can provide admirable precision for dated events, but disjunctions from stockpiling or re-use of timbers and loss of outer rings, and disparities from later repairs, are common problems for accuracy
- Most physics-based chronometric methods employ a sort of "clock" that most often comes from exponential decay of radioactive isotopes, but can also come from other processes, such as diffusion of water into obsidian
- All such methods require a kind of event that resets the "clock," such as the death of an organism (radiocarbon), heating of quartz crystals (thermoluminescence), cooling of volcanic lava (argon), or flaking of obsidian
- Radiocarbon dating entails several potential sources of bias, and the calibration process accounts for most of these. Remaining errors include precision related to statistical and other effects on measurement, and bias resulting from the selection of samples or discrepancies due to poor connection between the dated and target events.
- Bayesian modeling, when used with appropriate caution, provides excellent tools for refining the chronology of stratified series, for estimating the durations of site occupations, for testing the chronological relationships among different contexts, for tracking the dispersion of a technology or culture over large regions, and for analyzing large data sets.

References Cited

Ahlstrom, R. V. N., Dean, J. S., & Robinson, W. J. (1991). Evaluating tree-ring interpretations at Walpi Pueblo, Arizona. *American Antiquity, 56*(4), 628–644.

Aitken, M. J. (1990). *Science-based dating in archaeology*. New York: Longman.

Attenbrow, V., & Hiscock, P. (2015). Dates and demography: Are radiometric dates a robust proxy for long-term prehistoric demographic change? *Archaeology in Oceania, 50*(S1), 29–35.

Bamforth, D. B., & Grund, B. (2012). Radiocarbon calibration curves, summed probability distributions, and early Paleoindian population trends in North America. *Journal of Archaeological Science, 39*(6), 1768–1774.

Banning, E. B. (1997). *A better alternative to the "Mean ceramic date."* Paper presented at the Society for American Archaeology, Nashville, TN, 5 April 1997.

Banning, E. B., & Pavlish, L. A. (1978). Direct detection in radiocarbon dating. *Journal of Field Archaeology, 5*(4), 480–483.

Bayliss, A. (2009). Rolling out revolution: Using radiocarbon dating in archaeology. *Radiocarbon, 51*(1), 123–147.

Bayliss, A. (2015). Quality in Bayesian chronological models in archaeology. *World Archaeology, 47*, 677–700.

Bayliss, A., & Orton, C. (1994). Strategic considerations in dating, or 'how many dates do I need?'. *Institute of Archaeology Bulletin, 31*, 151–167.

Bayliss, A., van der Plicht, J., Bronk Ramsey, C., McCormac, G., Healy, F., & Whittle, A. (2011). Towards generational time-scales: The quantitative interpretation of archaeological chronologies. In A. Whittle, F. Healy, & A. Bayliss (Eds.), *Gathering time: Dating the early Neolithic enclosures of Southern Britain and Ireland* (pp. 17–59). Oxford: Oxbow Books.

Beckwith, S. J. (2017). *The built environment: Spatial and social interaction at Pot Creek Pueblo, New Mexico*. PhD dissertation. University of Toronto. Ann Arbor: ProQuest.

Bennett, C. L., Beukens, R., Clover, M., Gove, H., Liebert, R., Litherland, A. E., et al. (1977). Radiocarbon dating using electrostatic accelerators: Negative ions provide the key. *Science, 198*(4316), 508–510.

Bevan, A., Colledge, S., Fuller, D., Fyfe, R., Shennan, S., & Stevens, C. (2017). Holocene fluctuations in human population demonstrate repeated links to food production and climate. *Proceedings of the National Academy of Sciences, 114*(49) E10524-E10531.

Bishop, R. R. (2015). Summed radiocarbon probability distributions from cereal grains: Arable cultivation proxy or the 'archaeology of us' (a reply to Stevens and Fuller 2015). *World Archaeology, 47*(5), 876–881.

Boaretto, E. (2008). Determining the chronology of an archaeological site using radiocarbon: Minimizing uncertainty. *Israel Journal of Earth Sciences, 56*, 207–216.

Bowman, S. (1990). *Radiocarbon dating*. Los Angeles: University of California Press.

Brock, F., Higham, T. F. G., Ditchfield, P., & Bronk Ramsey, C. (2010). Current pretreatment methods for AMS radiocarbon dating at the Oxford Radiocarbon Accelerator Unit (ORAU). *Radiocarbon, 52*(1), 103–112.

Bronk Ramsey, C.B. (1998). *OxCal program*. http://info.ox.a.uk/departments/rlaha/xal/oxcal_h.html

Bronk Ramsey, C. B. (2008). Deposition models for chronological records. *Quaternary Science Reviews, 27*, 42–60.

Bronk Ramsey, C. B. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon, 51*(1), 337–360.

Bronk Ramsey, C. B. (2017). Methods for summarizing radiocarbon datasets. *Radiocarbon, 59*, 1809–1833.

Bronk Ramsey, C. B. (2018). OxCal 4.3. https://c14.arch.ox.ac.uk/oxcal.html. Accessed 4 Jan 2019.

Bronk Ramsey, C. B., van der Plicht, J., & Weininger, B. (2001). "Wiggle matching" radiocarbon dates. *Radiocarbon, 43*(2A), 381–389.

Brothwell, D. R., & Pollard, A. M. (Eds.). (2008). *Handbook of archaeological sciences*. New York: Wiley.

Buck, C. E., & Meson, B. (2015). On being a good Bayesian. *World Archaeology, 47*, 567–584.

Buck, C. E., Cavanagh, W. G., & Litton, C. D. (1996). *Bayesian approach to interpreting archaeological data*. New York: Wiley.

Buck, C. E., Christen, J. A., & James, G. N. (1999). BCal: An on-line Bayesian radiocarbon calibration tool. *Internet Archaeology, 7*. https://doi.org/10.11141/ia.7.1.

Christen, J. A. (1994). Summarizing a set of radiocarbon determinations: A robust approach. *Journal of the Royal Statistical Society C (Applied Statistics), 43*(3), 489–503.

Collis, J. (1971). Thoughts on radiocarbon dating. *Antiquity, 45*, 200–201.

Contreras, D. A., & Meadows, J. (2014). Summed radiocarbon calibrations as a population proxy: A critical evaluation using a realistic simulation approach. *Journal of Archaeological Science, 52*, 591–608.

Cook, G. T., Ascough, P. L., Bonsall, C., Hamilton, W. D., Russell, N., Sayle, K. L., et al. (2015). Best practice methodology for ^{14}C calibration of marine and mixed terrestrial/marine samples. *Quaternary Geochronology, 27*, 164–171.

Crema, E. R., Bevan, A., & Shennan, S. (2017). Spatio-temporal approaches to archaeological radiocarbon dates. *Journal of Archaeological Science, 87*, 1–9.

Crown, P. L. (1991). Evaluating the construction sequence and population of Pot Creek Pueblo, northern New Mexico. *American Antiquity, 56*(2), 291–314.

Dean, J. S. (1978). Independent dating in archaeological analysis. *Advances in Archaeological Method and Theory, 1*, 223–255.

Dunbar, E., Cook, G. T., Murdoch, I., Xu, S., & Fabel, D. (2018). Identification of fraudulent-age whiskies using accelerator mass spectrometry (AMS) radiocarbon (^{14}C) analyses. In *Proceedings of the worldwide distilled spirits conference 2017*. Glasgow: Scottish Universities Environmental Research Centre. http://eprints.gla.ac.uk/168230/

Dye, T. S. (2011). A model-based age estimate for Polynesian colonization of Hawa'i. *Archaeology in Oceania, 46*(3), 130–138.

Dye, T. S., & Buck, C. E. (2015). Archaeological sequence diagrams and Bayesian chronological models. *Journal of Archaeological Science, 63*, 84–93.

Fitzpatrick, S. M. (2006). A critical approach to using ^{14}C dating in the Caribbean: Using chronometric hygiene to evaluate chronological control and prehistoric settlement. *Latin American Antiquity, 17*(4), 389–418.

Freeman, J., Byers, D. A., Robinson, E., & Kelly, R. L. (2018). Culture process and the interpretation of radiocarbon data. *Radiocarbon, 60*(2), 453–467.

Geyh, M., & Schlüchter, C. (1998). Calibration of the ^{14}C time scale beyond 22,000 BP. *Radiocarbon, 40*, 475–482.

Hamilton, W. D., & Krus, A. M. (2017). The myths and realities of Bayesian chronological modeling revealed. *American Antiquity, 83*(2), 187–203.

Hua, Q., Barbetti, M., & Rakowski, A. Z. (2013). Atmospheric radiocarbon for the period 1950–2010. *Radiocarbon, 55*(4), 2059–2072.

Jones, N. (2020). Carbon dating, the archaeological workhorse, is getting a major reboot. *Nature, 581*. https://doi.org/10.1038/d41586-020-01499-y.

Jones, M., & Nicholls, G. (2002). New radiocarbon calibration software. *Radiocarbon, 44*, 663–674.

Jöris, O., & Weninger, B. (1998). Extension of the ^{14}C calibration curve to ca. 40,000 cal BC by synchronizing Greenland ^{18}O/^{16}O ice core record and North Atlantic foraminifera profiles: A comparison with U/Th coral data. *Radiocarbon, 40*, 495–504.

Kitagawa, H., & van der Plicht, J. (1998). A 40,000-year varve chronology from Lake Suigetsu, Japan: Extension of the ^{14}C calibration curve. *Radiocarbon, 40*, 505–515.

Lamont-Doherty Earth Observatory, Columbia University. (n.d.) Tree Ring Lab software. https://www.ldeo.columbia.edu/tree-ring-laboratory/resources/software. Accessed 18 Dec 2018.

Lamoureux, S. (2002). Varve chronology techniques. In W. M. Last & J. P. Smol (Eds.), *Tracking environmental change using lake sediments* (pp. 247–260). Dordrecht: Springer.

Lanos, P., Philippe, A., Lanos, H., & Dufresne, P. (2016). Chronomodel: Chronological modelling of archaeological data using Bayesian statistics (version 1.5). http://www.chronomodel.fr

Libby, W. F. (1955). *Radiocarbon dating* (2nd ed.). Chicago: University of Chicago Press.

Litton, C. D., & Zainodin, H. J. (1991). Statistical models of dendrochronology. *Journal of Archaeological Science, 18*, 429–440.

Lockyear, K. (2012). Dating coins, dating with coins. *Oxford Journal of Archaeology, 31*(2), 191–211.

Lorrain, D. (1968). An archaeologist's guide to nineteenth-century American glass. *Historical Archaeology, 2*, 35–44.

Malainey, M. E. (2011). *A consumer's guide to archaeological science*. New York: Springer.

Michczynska, D. J., & Pazdur, A. (2004). Shape analysis of cumulative probability density function of radiocarbon dates set in the study of climate change in the Late Glacial and Holocene. *Radiocrbon, 46*(2), 733–744.

Millard, A. R. (2006). Bayesian analysis of Pleistocene chronometric methods. *Archaeometry, 48*(2), 359–375.

Nayton, G. (1992). Applying frontier theory to a western Australian site: The problem of chronological control. *Australasian Historical Archaeology, 10*, 75–91.

Ortman, S. G. (2016). Uniform probability density analysis and population history in the northern Rio Grande. *Journal of Archaeological Method and Theory, 23*, 95–126.

Orton, C. (1980). *Mathematics in archaeology*. Cambridge: Cambridge University Press.

Reimer, P. J., Bard, E., Bayliss, A., et al. (2013). IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon, 55*(4), 1869–1887.

Roberts, J. M., Mills, B. J., Clark, J. J., Haas, W. R., Huntley, D. L., & Trowbridge, M. A. (2012). A method for chronological apportioning of ceramic assemblages. *Journal of Archaeological Science, 39*, 1513–1520.

Robinson, E., Zahid, H., Codding, B. F., Haas, R., & Kelly, R. L. (2019). Spatiotemporal dynamics of prehistoric human population growth: Radiocarbon 'dates as data' and population ecology models. *Journal of Archaeological Science, 101*, 63–71.

Samford, P. M. (1997). Response to a market: Dating English underglaze transfer-printed wares. *Historical Archaeology, 31*(2), 1–30.

Shennan, S., Downey, S. S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., et al. (2013). Regional population collapse followed initial agriculture booms in mid-Holocene Europe. *Nature Communications, 4*, 2486. https://doi.org/10.1038/ncomms3486.

Smith, M. (2016). The use of summed-probability plots of radiocarbon data in archaeology. *Archaeology in Oceania, 51*(3), 214–215.

South, S. (1977). *Method and theory in historical archaeology*. New York: Academic.

Speer, J. H. (2010). *Fundamentals of tree-ring research*. Tucson: University of Arizona Press.

Spriggs, M. (1989). The dating of the island S.E. Asian Neolithic: An attempt at chronometric hygiene and linguistic correlation. *Antiquity, 63*, 587–613.

Steponaitis, V., & Kintigh, K. (1993). Estimating site occupation spans from dated artifact types: Some new approaches. In J. B. Stoltman (Ed.), *Archaeology of Eastern North America: Essays in honor of Stephen Williams* (pp. 349–361). Jackson: Mississippi Department of Archives and History.

Stuiver, M., Reimer, P. J., & Reimer, R. (2019). Calib radiocarbon calibration, ver. 7.1. http://calib.org/calib/. Accessed 4 Jan 2019.

Surovell, T. A., & Brantingham, P. J. (2007). A note on the use of temporal frequency distributions in studies of prehistoric demography. *Journal of Archaeological Science, 34*, 1868–1877.

Surovell, T. A., Byrd Finley, J., Smith, G. M., Brantingham, P. J., & Kelly, R. (2009). Correcting temporal frequency distributions for taphonomic bias. *Journal of Archaeological Science, 36*, 1715–1724.

Taylor, R. E. (2014a). Major anomalies. In R. E. Taylor & O. Bar-Yosef (Eds.), *Radiocarbon dating: An archaeological perspective* (2nd ed., pp. 43–64). Walnut Creek: Left Coast Press.

Taylor, R. E. (2014b). Samples and sample pretreatment. In R. E. Taylor & O. Bar-Yosef (Eds.), *Radiocarbon dating: An archaeological perspective* (2nd ed., pp. 65–97). Walnut Creek: Left Coast Press.

Taylor, R. E. (2014c). Critical evaluation of radiocarbon data. In R. E. Taylor & O. Bar-Yosef (Eds.), *Radiocarbon dating: An archaeological perspective* (2nd ed., pp. 130–171). Walnut Creek: Left Coast Press.

Taylor, R. E., & Aitken, M. J. (Eds.). (1997). *Chronometric dating in archaeology*. New York: Springer.

Taylor, R. E., & Bar-Yosef, O. (2014). *Radiocarbon dating: An archaeological perspective* (2nd ed.). Walnut Creek: Left Coast Press.

Thompson, V. D., Jefferies, R. W., & Moore, C. R. (2018). The case for radiocarbon dating and Bayesian analysis in historical archaeology. *Historical Archaeology, 53*(1), 181–192.

Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Manning, K., et al. (2014). Reconstructing regional population fluctuations in the European Neolithic using radiocarbon dates: A new case-study using an improved method. *Journal of Archaeological Science, 52*, 549–557.

Wesler, K. (2014). Assessing precision in formula dating. *Historical Archaeology, 48*(2), 173–181.

Williams, A. N. (2012). The use of summed radiocarbon probability distributions in archaeology: A review of methods. *Journal of Archaeological Science, 39*(3), 578–589.

21 Archaeological Illustration and Publication

Much more than just a medium of representation, visual image creation—above all, drawing—constitutes a multi-sensory means of both exploration and interpretation. . . . Image creation in general, I suggest, and the process of drawing in particular, comprise fundamental archaeological experiences that all serious students of the discipline should be required to practise as a basic element of their training.

James (2015: 1200)

Lithic drawing taps the illustrator's expert judgement, ensuring that the system tracks the determinables needed to test hypotheses for which the stones are evidence.. . . This is why lithic illustrators must know some archaeology and why they are often closely associated with archaeological labs.

Lopes (2018)

Archaeological illustrations are not just pictures. As noted in Chap. 4, they involve graphical information language—a formal, visual way to describe artifacts and other archaeological information in a consistent way (Addington 1986: ix–xiv; Shirvaklar 2017: 219). That is why we still use them, even in an age of digital photographs, RTI (reflectance transformation imaging), and 3D scanning. Drawings of lithics, pottery, plans, and stratigraphic sections are not realistic or artistic renderings of what the artist sees, but technical drawings that present a selection of information that some archaeologists consider important, while omitting an infinite number of details considered less important, less relevant, or distracting (Adkins and Adkins 1989: 5–6). That is not to say that the selection is completely arbitrary; archaeologists employ a wide range of conventions for the minimal information that they expect to find in an archaeological illustration. For some of these, such as those for lithics, the conventions are well developed. For others, there is less agreement. Either way, it is important for archaeological publications to include clear keys that explain what these conventions are.

While we use computer software to generate most illustrations today (e.g., Graser and Peterson 2018), at least in their final versions, it is still important for archaeologists to have at least some facility with drawing illustrations by hand, if only as a draft version to scan and finish with graphics software. You do not need to be a great artist to produce a useful archaeological illustration, but you do need to have patience and learn some basic skills (Dillon 1985; Griffiths et al. 2007; Steiner 2005).

In addition, it is useful to have some understanding of the process by which archaeological interpretations, through text and illustrations, become publications, in print, online or both.

21.1 Early Archaeological Illustration

Early illustrations by European and Asian antiquarians were less conventional and more "realistic" than is usual today, with perhaps more emphasis on aesthetics (Figs. 21.1 and 21.2). Nonetheless, many of these provide useful evidence for artifacts or monuments that have since disappeared or suffered damage, and we can sometimes even recognize attributes in these images that the illustrators themselves might not consciously have considered important. In addition, some early illustrations already begin to show conventions that we would recognize today, such as scales or cutaway views.

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3_21

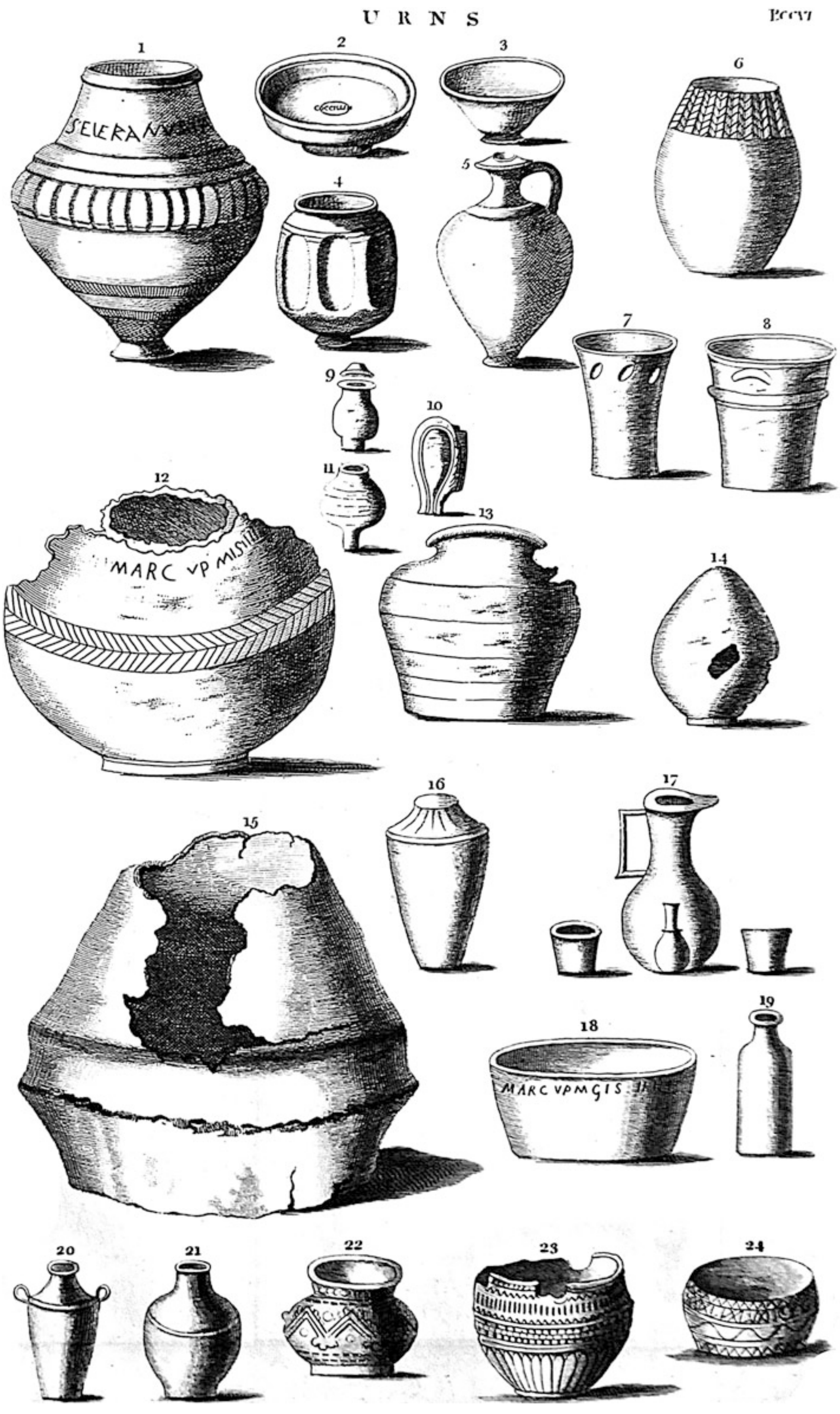


Fig. 21.1 Early nineteenth-century plate of vessels from British and Irish archaeological sites (Cambden 1806: plate 206)

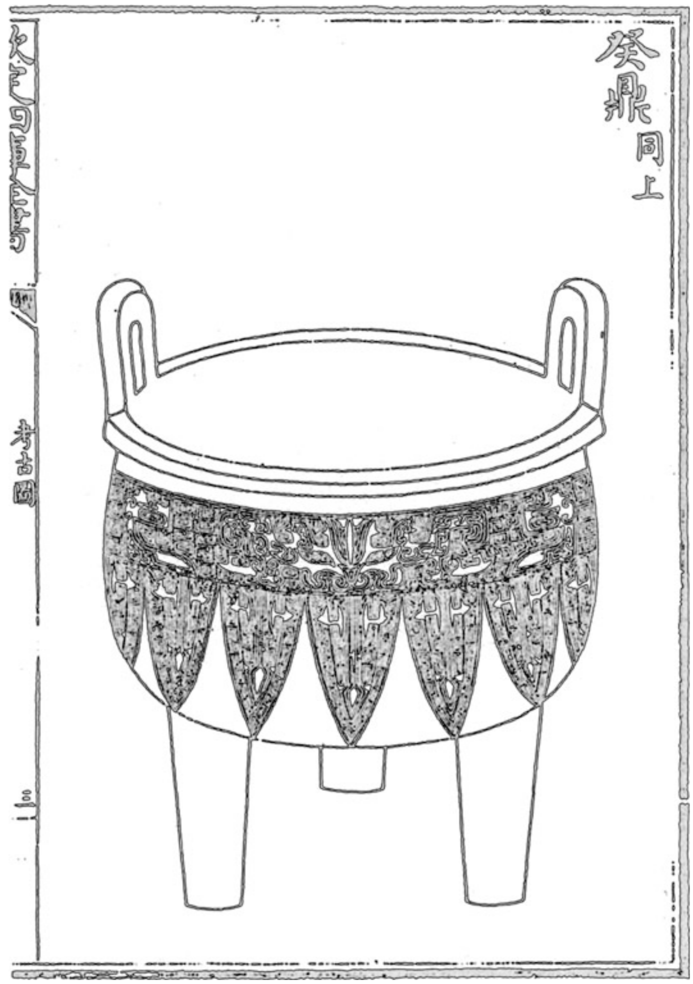


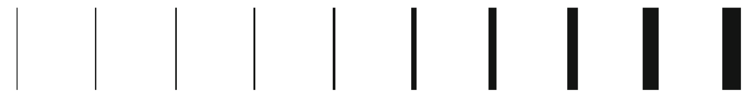
Fig. 21.2 Decorated *ding* (cauldron), probably of the Han dynasty, as depicted in a Chinese antiquarian work (Lü Dalin 1092: 1.3a)

21.2 Style of Representation and Basic Conventions

Archaeological illustrators usually adhere to sets of conventions or "style," although different projects and archaeological traditions vary somewhat in these styles.

By "style" we mean the combination of views, line thicknesses, conventions for shading or representing materials, and other conventions that will allow consistent interpretation of the images. One of the first things you should do if you are about to illustrate artifacts, plans, or sections for a project is to decide what styles you will use. Consistency in style makes the published illustrations, indeed the whole publication, look much better and, more importantly, facilitates comparison among images. The style you choose encodes information that you want to convey to an audience. For example, stippling might indicate cortex on a stone tool, hatching with a particular line thickness and spacing might indicate post-depositional damage, while another style of hatching might indicate red slip on pottery. Reconstructed or missing parts of an artifact might be indicated with dashed, dotted or grayed lines, while the code for animal burrows on a section drawing from an excavation might be a pattern of small, random line segments. All these drawing conventions are parts of an information language.

To make it easy for you or other illustrators to follow the conventions for your style, you should make a manual or "style sheet" that documents them all and, for certain symbols or patterns that you use regularly, such as North arrows and scales, it is useful to make up templates for them in a variety of sizes so that you can easily "paste" them into your finished illustrations. When you decide on line weights (thicknesses) and hatching for these styles, you should keep in mind whether they will be reproduced at the same scale as drawn,



Point Size	0000	000	00	0	1	2	3	4	6	7
Size in mm	0.18	0.25	0.30	0.35	0.50	0.60	0.80	1.0	1.4	2.0

Fig. 21.3 Table of point sizes and line weights for technical pens and illustration software

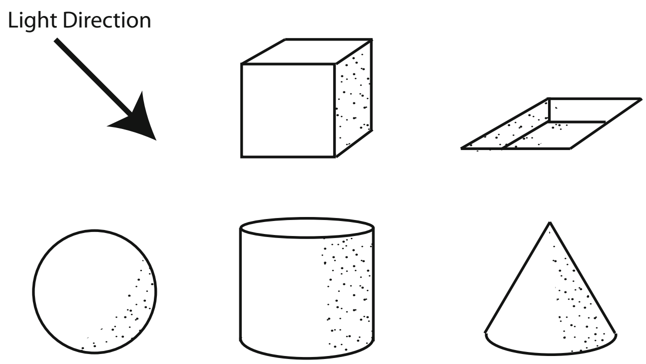


Fig. 21.4 Shading objects to simulate illumination from the upper left. (Modified from Newlands and Breede 1976: 132)

or be reduced in size for publication, possibly down to 25% or even 12.5% of the original size. Otherwise, in the published version, thin lines could disappear or closely-spaced hatching could bleed together. To check on the effects of scale reductions, it is useful to make a table of point sizes (Fig. 21.3) for line weights and fonts and any important symbols you plan to use and then reduce this test sheet to a variety of scales. Post these test sheets where they are visible to illustrators.

By convention, archaeologists depict artifacts as though the light is coming from upper left, at an angle close to 45° (Fig. 21.4). The easiest way to do this is to light it that way while you are drawing it, but experienced illustrators can shade drawings this way no matter what their lighting. This ensures that all the artifacts depicted on a single page seem to be illuminated in a consistent way, facilitating comparison and making the page look more unified.

Most styles for artifact illustrations require multiple views, such as the dorsal and ventral sides of a lithic flake and at least one edge. In that case, you should use standard orthographic projection to arrange these views so that they can be "folded" into correct position (Fig. 21.5). Each view shows the surface nearest to it in an adjacent view, but at an angle that differs by 90°, much like the sides of a box.

For maps and plans, it is usually best to keep North more-or-less at the top unless you have strong reason to do otherwise, such as orienting plans of buildings consistently. In the latter cases, including North arrows in the image may be important, or not, depending on the illustration's purpose. Normally, however, every map or plan should have a North arrow (or, for polar regions, lines of longitude), and some indication of scale that will still be useful if the publisher or user reduces or enlarges the image. Saying the scale is 1:50, for example, is worthless if the image might be rescaled before printing, or if it will be published digitally on a web site or in a PDF that users will scale as they see fit.

21.3 Basic Equipment and Supplies

Even when most of your illustration work is on a computer, you will need some basic drawing equipment. Among these are a high-intensity lamp (not a fluorescent light) to illuminate details on artifacts and accentuate shadows, and a magnifying glass to help you examine small details. If possible, use a magnifier with a large lens (~8 cm), mounted on a base or swinging arm that allows you to position it while keeping your hands free. Other tools that most artifact illustrators need are calipers for measuring artifacts, preferably with plastic "jaws" that will not damage the artifacts' edges, a "forma-gauge" or "profile gauge" that allows you to reproduce complex shapes (Fig. 21.6), metal straight-edges, preferably with bevelled edges or raised on thin cork bases, and technical pencils and leads. Pencils that accept 0.5 mm 5H leads work well for most applications.

While drawing pens are not as essential as they once were, because archaeological illustrators mainly "ink" their work digitally instead of on paper or Mylar™, inking lithic illustrations usually is, even today, best done by hand with pen and ink. It is certainly possible to complete these illustrations on a computer, as many archaeological projects now do, but freehand inking by a skilled illustrator usually results in a more compelling illustration that shows the ripples in the flake scars very effectively (Figs. 11.8 and 11.9), and scanning these creates digital files. Computer-drawn attempts to represent the ripples tend to be only conventional and not as effective at portraying the depth of flake scars unless the illustrator invests a great deal of time in each image.

Consequently, there can still be reasons to invest in pens and ink, even when most final illustrations are digital. Technical drawing pens typically come in sets with a variety of point sizes from 000 or 00 up to 3 or 4 (0.18–1.0 mm) and have refillable ink cartridges. Most artifact illustrations are substantially reduced in scale during publication, in which

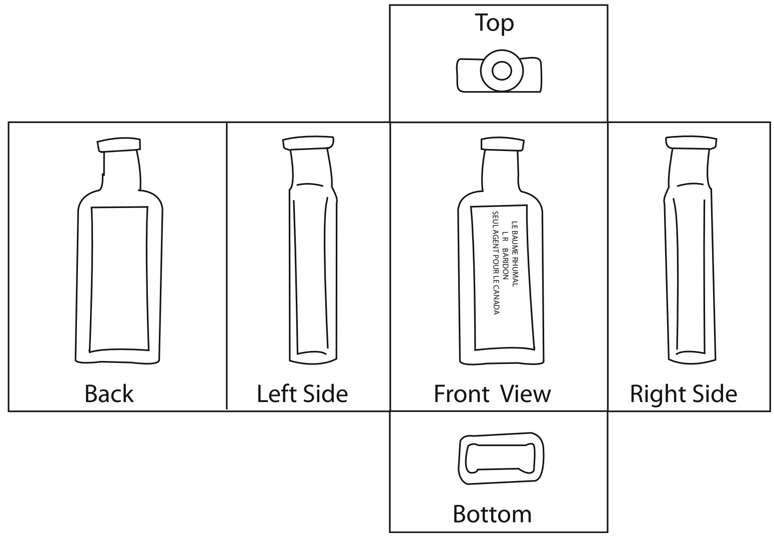


Fig. 21.5 Orthographic projection of a glass pharmaceutical bottle (W. Wadsworth)

Fig. 21.6 Use of a profile gauge to reproduce complex shapes

case pen sizes of 1–4 are necessary. Some lithic illustrators also like to use "crow-quill" or calligraphy pens to draw the ripples; their nibs have the advantage of allowing the illustrator to vary the width of lines, so that they gradually thin towards the end of each stroke. Pens need careful cleaning, especially if left unused for several days, to prevent them from clogging. Cleaning technical pens requires care not to damage the tiny wire that transmits the ink. For hand-inked drawings, sprinkling draftsman's powder (powdered rubber eraser) helps to keep the drawing surface dry and prevent smearing.

Handy tools for drawing by hand are right-angle triangles, flexible rulers or "French curves" for tracing curved things like contour lines, guillotine paper cutter, scissors, box-cutters, and cutting pads that protect counter surfaces when slicing through paper or Mylar. It is also convenient to have a "nonphoto blue" or "drop-out blue" pencil. This allows illustrators to write instructions or labels on drawings that will not show up when they are photographed or scanned. Almost any kind of paper or vellum will do for pencil drawings for immediate scanning, but paper shrinks or expands during storage with variations in humidity, so the scale of drawings on paper is not trustworthy after several months. For longer-term storage, especially of inked drawings, Mylar is much better—it is more environmentally stable and very durable—and large drawings can be stored flat in a map cabinet.

When it is necessary to do a lot of drawing prior to finishing illustrations on a computer, it is beneficial to use a drafting table or light table and stool at ergonomically comfortable heights, placed near a window to ensure good, natural light. These used to be standard in archaeological labs but have become much less common as archaeologists came to rely on computers for most aspects of illustration. However, drafting tables can be adjusted to various angles so that illustrators do not have to lean over and risk back strain as they would while drawing on a horizontal surface, while light tables make it easy to trace images through most kinds of paper and vellum.

Today, the most essential illustration equipment includes a computer with a large monitor, a flat-bed scanner with at least 300 dpi resolution or large digitizing tablet, and software that facilitates making vector-based graphics from the scanned images. Some labs may have 3D scanners that have definite advantages (see below). It is also essential to have a printer on which to test the output to check on things like line sizes and legibility, and sufficient storage space, whether local or cloud storage, for the many large files that typically result from illustration work. Several very good graphics, illustration and CAD software platforms are available but vary in their features, so it is necessary to do research to ensure that the ones selected meet the project's needs. There are also specialized software packages for creating Harris matrices, statistical graphs, and contour maps from total station data or photogrammetry. It is best to select ones that allow users to save images in file formats that other illustration or CAD software can read. This makes it possible to edit them to fit a project's or publisher's style and override the software packages' own styles and defaults.

21.4 Lithic Illustration

Lithic illustrators usually take time to become proficient and the best ones are usually also lithic analysts, rather than professional artists, because good lithic illustration requires fairly thorough understanding of lithic technology (see Chap. 11; Addington 1986; Chase 1985; Raczynski-Henk 2017).

The first step in lithic illustration is to orient the piece correctly. For flakes and blades, the proximal end, where the bulb and platform would occur, is at the bottom. Bifaces (handaxes), projectile points, drills, awls, and other pointed tools are oriented with the pointed end upwards, and endscrapers with the retouched end at the top (Addington 1986: 43–48). Single-platform cores are oriented with the platform at the top, while choppers are shown with the cutting edge at the top, and burins such that the burin blows (marked by small arrows) point predominantly downward.

Most conventions call for multiple views in orthographic projection (Figs. 21.5 and 21.7). If drawing these on paper, you should draw them on the same sheet, positioned

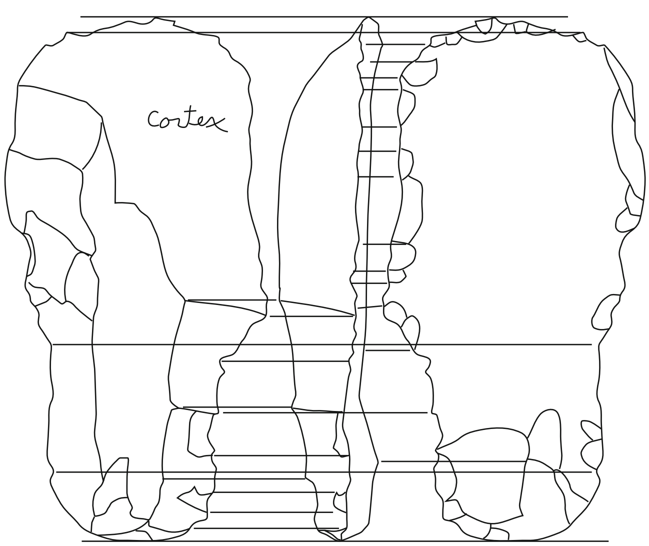


Fig. 21.7 A "selective grid" to ensure that features appearing in two or three adjacent views in a preliminary drawing of a lithic artifact are correctly aligned

correctly, so that you can use vertical and horizontal lines (in "drop out blue" if you will later scan these) to ensure that "landmarks" that show in two or more views are aligned correctly. If you plan to do the whole illustration on a computer, based on scans or photographs of the lithic itself, you could have some distortion from **parallax** that you will have to correct with caliper measurements. Using a grid on the screen will allow you to align those landmarks correctly and ensure that each view is at the same scale.

Whether with pencil or digital cursor, you begin by drawing the outline of the artifact in either dorsal or ventral view, using caliper measurements to make sure that distances as portrayed in the drawing are very close to the corresponding distances on the artifact. If you try to trace an artifact directly, use a very sharp pencil and try to hold the pencil in such a way that the point is very close to the artifact's edge, to avoid parallax, and keep in mind that when you trace the artifact later on, digitally or not, you should stay *inside* these lines. However, tracing artifacts directly is not recommended, as it could damage their edges. When the outline is complete to your satisfaction, you can either fold the paper and trace the outline on a light table or, for digital tracing, copy the outline and then flip it in the software. Thus, you have two identical but mirror-images that you can use as outlines for the ventral and dorsal views.

When drawing the dorsal view, pay close attention to the direction from which flakes were struck off, as you will need to depict ripples as though they radiate from the point of percussion. On illustrations of fine-grained materials, most conventions call for these ripples to be smooth and perhaps tapering; on ones of coarse materials, such as quartz or quartzite, or weathered pieces, some conventions show discontinuous rippling to give an impression of a rougher surface (Fig. 21.8). If there is any cortex on the dorsal side, some styles will distinguish this with a kind of stippling (random dots). Where possible, it is helpful to show the sequence of flake removals (Addington 1986: 14).

Taking the other outline, you will need to represent the bulb of percussion, if present, and indicate broad ripples across the ventral surface that extend outwards from the point of percussion. Keep in mind the lighting from upper left as you vary the density of ripples, especially on and near the bulb, to give an impression of how prominent the bulb is. If they are visible, be sure to include details of the platform remnant and fissures at the bulb. There may also be small flake scars around the edges from retouch or damage that you will draw using the same techniques you used for the dorsal side.

Particularly when the edges are retouched, it is common to add side or end views. Again, use parallel lines or a grid to align landmarks in these views with corresponding ones in the dorsal or ventral view, and use calipers to make sure that you portray thicknesses at these landmarks correctly.

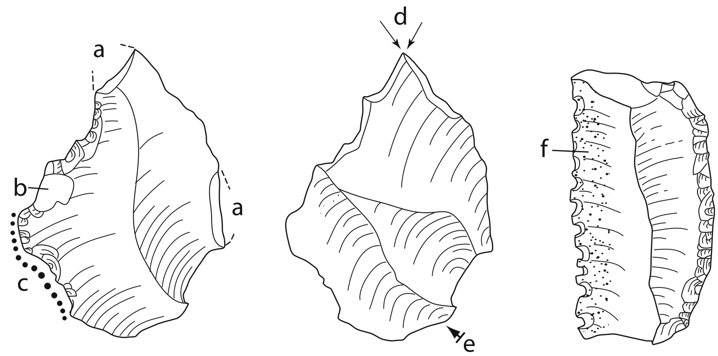


Fig. 21.8 Some conventions for broken (**a**), "bag-retouched" (**b**), and abraded areas (**c**), burin blows (**d**), atypical position of bulb (**e**), and polish (**f**) (**a**–**e** after Addington 1986: 21–22)

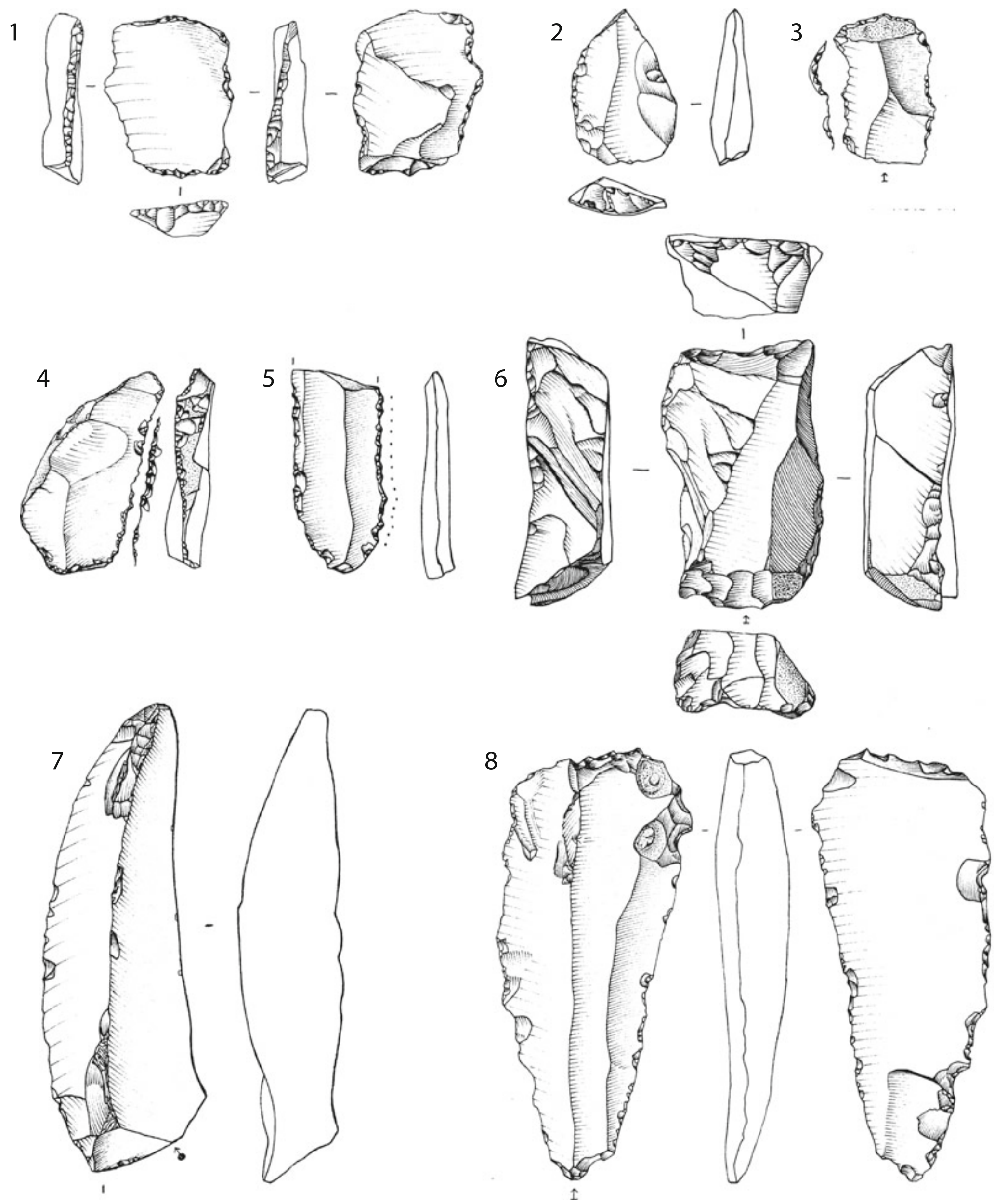


Fig. 21.9 Example of a lithic plate with multiple views and correct orientation for blades (three with "sickle sheen"), points, bifacial knife, and axes (J. Pfaff)

A variety of common conventions indicate edge details and other features. For example, on burins, we use small arrows to indicate the position and direction of burin blows (Fig. 21.8). Where there are several, closely-spaced burin blows, staggering the arrows reduces crowding and makes the image clearer. Barred arrows (with a "T" shape at their base) serve to indicate the position of the bulb of percussion and direction of flake detachment whenever those are not in the "usual" position at the bottom of the drawing. This usually happens on pointed tools that are oriented point-up.

Some illustrators use dashed or solid line segments between views in the orthographic projection to guide viewers' eyes from one view to another, although this is not strictly necessary. However, if you also add a cross-sectional view (Fig. 21.9) short line segments are important for indicating the location of the "slice" through the artifact.

Some conventions include styles to indicate thermal damage, such as pot-lid fractures and fine cracks, while dashed lines may indicate missing parts of broken flakes or tools, and rows of large dots, graduated in size, may indicate abrasion. "Bag retouch"—flake scars that are likely due to post-excavation edge damage—may be indicated by omitting the ripples.

21.5 Illustrating Pottery and Stone, Glass or Metal Vessels

Early archaeological illustrations of pottery and other vessels typically depict them, whether whole or fragmentary, much as they would look to a person viewing them from the outside, or as they would appear in a photograph (Figs. 21.1 and 21.2). Today, however, it is much more common to follow conventions that allow us to depict the interior, exterior, and radial section, and sometimes other views as well, while also reconstructing as much of a whole vessel as we can infer from a fragment.

Thus, typical pottery illustrations show each sherd or vessel as though we have cut out one-quarter of a vessel's circumference (Fig. 21.10), and this is common for stone vessels as well (Shirvaklar 2017). Most European projects show the cut-out on the left, and most North American ones show it on the right; it does not matter, as long as you are consistent, unless you are publishing in a journal whose style dictates one of these choices (see "Journal styles" below). For sherds and incomplete vessels, the reconstruction extends either from the rim downwards as far as possible, or from the base upwards as far as the evidence allows. As most vessels are nearly circular in plan view, the illustrations typically depict them as though they are perfectly circular. When this is clearly unrealistic, as with vessels that have oval, rectangular or more complex shapes, top or bottom views may be necessary to make this clear.

However, drawing most vessels begins by establishing their inside or outside diameter, assuming that they are circular. For rims and bases, we can do this easily with a diameter chart (Fig. 12.4), with nested arcs that correspond with increments in diameter and radiating line segments that indicate percentage of whole circumference represented when "stancing" a sherd on the chart. For body sherds, one can estimate diameter by using a profile gauge against the interior of the sherd, taking care to hold the gauge as close to the plane of stance as possible, and then place the gauge against the diameter chart. Once diameter is determined, it becomes the length of a horizontal line segment—the stance line—drawn on paper or vellum or on a computer, and then the illustrator bisects it by a perpendicular line segment that extends downwards to mark the central axis of the vessel (Fig. 21.14a).

The next step is to draw the profile and radial section of the sherd or vessel. There are several ways to do this, including using right-angle triangles and calipers to trace the outline of large, whole vessels, but the following outlines two of the most common and reasonably fast approaches for sherds.

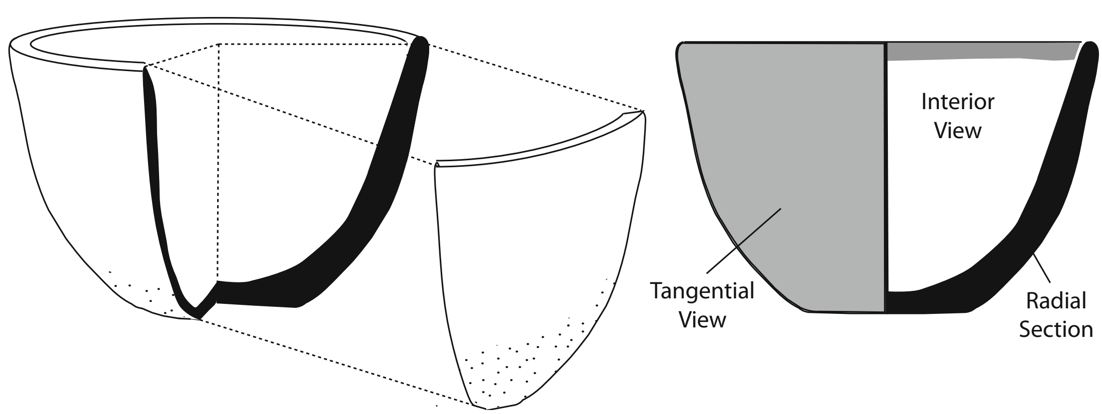


Fig. 21.10 Conventional vessel illustrations reconstruct as much of the vessel as possible and show a cut-away of a quarter-section to show the interior and radial section, as well as exterior or tangential view

One is to use the profile gauge, held as nearly vertically as possible (i.e., perpendicular to the vessel stance), to reproduce the interior and exterior profiles of a sherd (Fig. 21.6). The forma-gauge or profile gauge is a tool with a large number of thin, metal or plastic teeth sandwiched between metal brackets in a comb-like arrangement, and held somewhat loosely in place magnetically or by friction. Outside archaeology, these are used by plasterers and cabinet-makers to reproduce profiles of cornices and moldings, but they work just as well to reproduce complex vessel profiles, although plastic ones are better than metal to avoid damage to archaeological objects. When someone pushes the profile gauge's teeth against a complex surface, taking care that the gauge is perpendicular to the sherd's surface (i.e. parallel to a radial section), the teeth move to conform to its shape and retain an image of this shape when the gauge is removed. One can then place the gauge against paper or vellum and trace the shape with a sharp pencil. It is necessary to do this separately for the interior and exterior profiles, and sometimes rim details, and then piece them together, using caliper measurements to get the wall thickness right. It is important to orient the profile correctly relative to the stance line, which involves either stancing the sherd and measuring some angles with a goniometer or, somewhat better, stancing it against a table (upside-down for rims), and using right-angle triangles to find the horizontal, radial distance from the rim, and vertical distance from the table, to the farthest point along the break (Fig. 21.11).

Another method only works on sawn sherds. Sawing is not always desirable or possible, for reasons of conservation ethics (see Chap. 9) but is a necessary step to produce thin-sections for petrographic work (see p. 197), in which case an illustrator can take advantage of this. Whenever possible, it is best to have radial sections with a lapidary saw, taking care to select a place on a sherd that provides the maximum vertical profile or will intersect important features, such as handles (Fig. 21.12). This not only reveals important features for the petrography, it also provides the best possible section for illustrating the vessel.

Cut sherds make drawing much easier. One can begin by tracing the edge of a right-angle block to draw a stance line on paper, and leave the block in place to stance the cut sherd against it. With the cut section lying against the paper, push the sherd against the block until the rim (or base) is flush against it, meaning that it is in contact with the block along its whole perimeter (Fig. 21.13). The sherd is now "at stance" and it is straightforward to trace it with a sharp pencil held at a constant angle in such a way as to fit tightly against the sherd surface. Holliday (1976) recommends constructing a special holder for the pencil to make the tracing as tight to the profile as possible. Although it is a good idea to check the thickness of the tracing against some caliper measurements, generally this method reproduces profiles more accurately

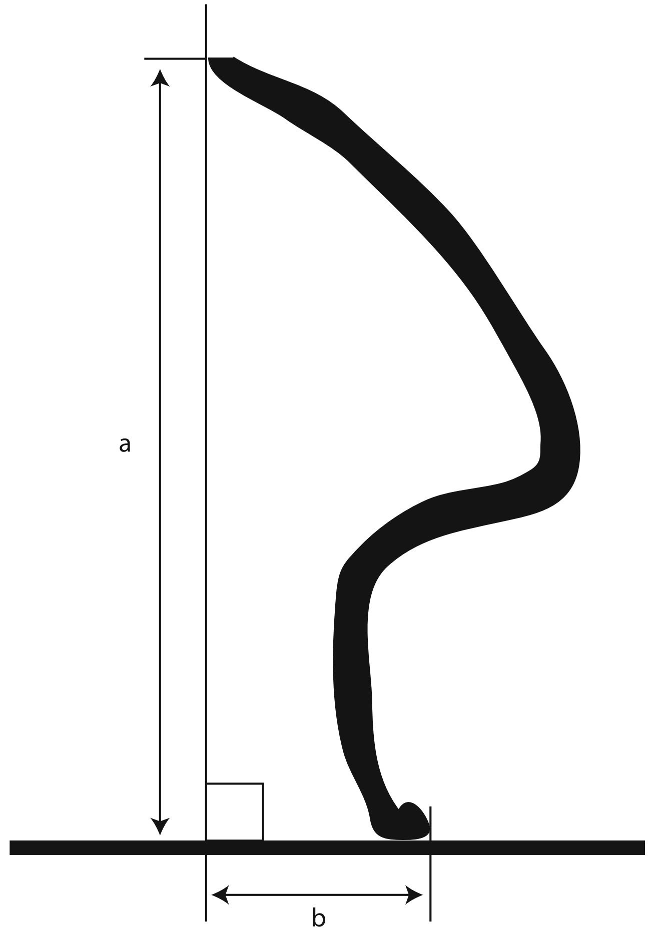


Fig. 21.11 Measuring the height and radial distance of a sherd that is stanced correctly to make sure the profile is oriented accurately

and more quickly than the profile-gauge method, but it is also destructive, since it requires sawing.

After tracing the outline of both interior and exterior surfaces, by either method, it is useful to make small tick marks to indicate places where one of the surfaces has horizontal grooves, decorative panels, or carinations (inflection points or sharp turns in the profile); these are reminders to add these features to the final drawing. Notations on the drawing indicate paint colors, highlight the extent of slip or burnishing, or any other observation relevant to the finished illustration, including any suspicions about the accuracy of the stance or diameter.

With the profile complete, the next step is to reflect it about the vessel's central axis (that vertical line centered on the stance line in Fig. 21.14c). If done on paper, that means flipping or folding the profile drawing and then tracing only the outer profile on the opposite end of the stance line, to create a mirror image. On a computer, you would scan the profile, trace its outline with the cursor, copy the resulting polygon, and then reflect the copy vertically and move it into position, then "cut" the polygon or delete the

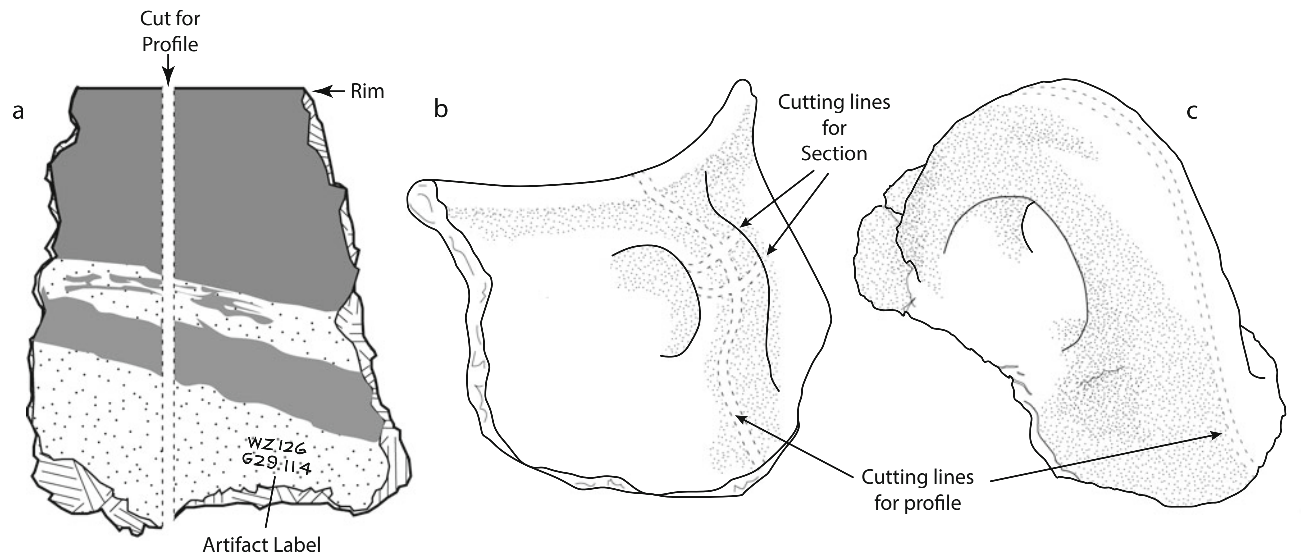


Fig. 21.12 Cuts through sherds should be on a radial plane in such a way as to obtain the longest vertical profile (**a**) or to intersect vertically important features, such as handles (**b** and **c**). (E. Banning and Y. Salama, after Holliday 1976)

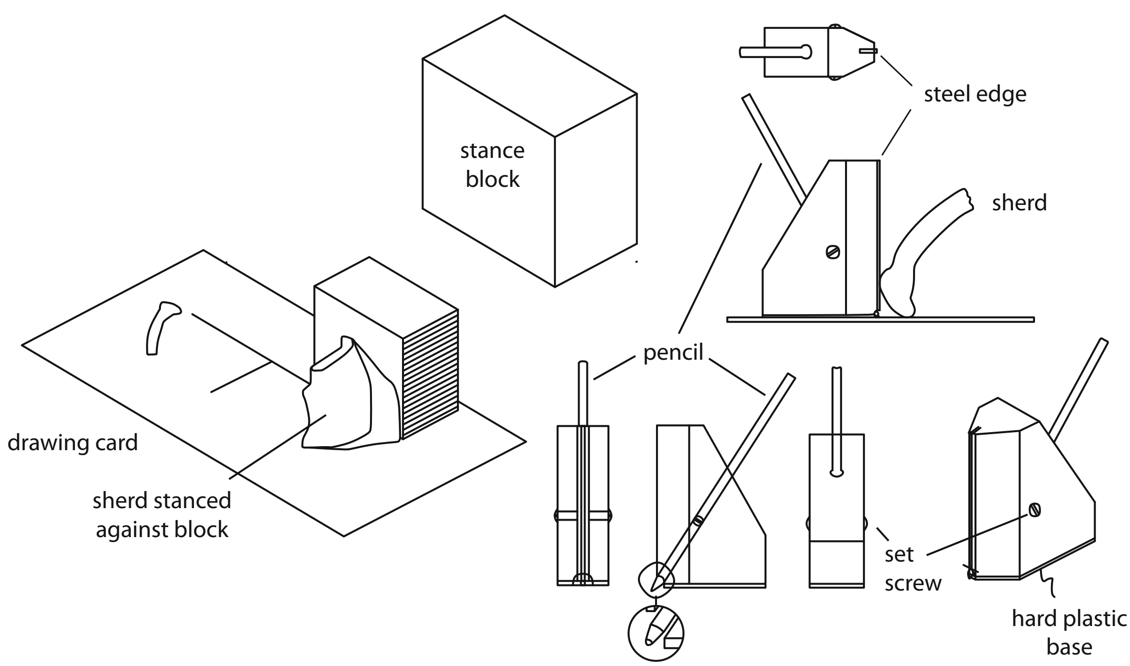


Fig. 21.13 Specifications for stance block and pencil scriber to outline a stanced rim sherd. (Y. Salama, after Holliday 1976)

trace of the inner surface so that only the exterior surface remains.

If the exterior is on the left and the cross-section on the right, the next step is to indicate any exterior features, such as grooves, carinations, knobs or decoration, on the left side, and interior ones on the right. Generally, the idea is to reconstruct as much of the original vessel as possible but, for the sake of honesty, it is common to use a convention, such as dashed lines, to reconstruct uncertain features, such as portions of a painted panel that are not preserved but are reasonably probable (Fig. 21.15), or to interrupt the decoration to indicate the missing portions. When drawing complex

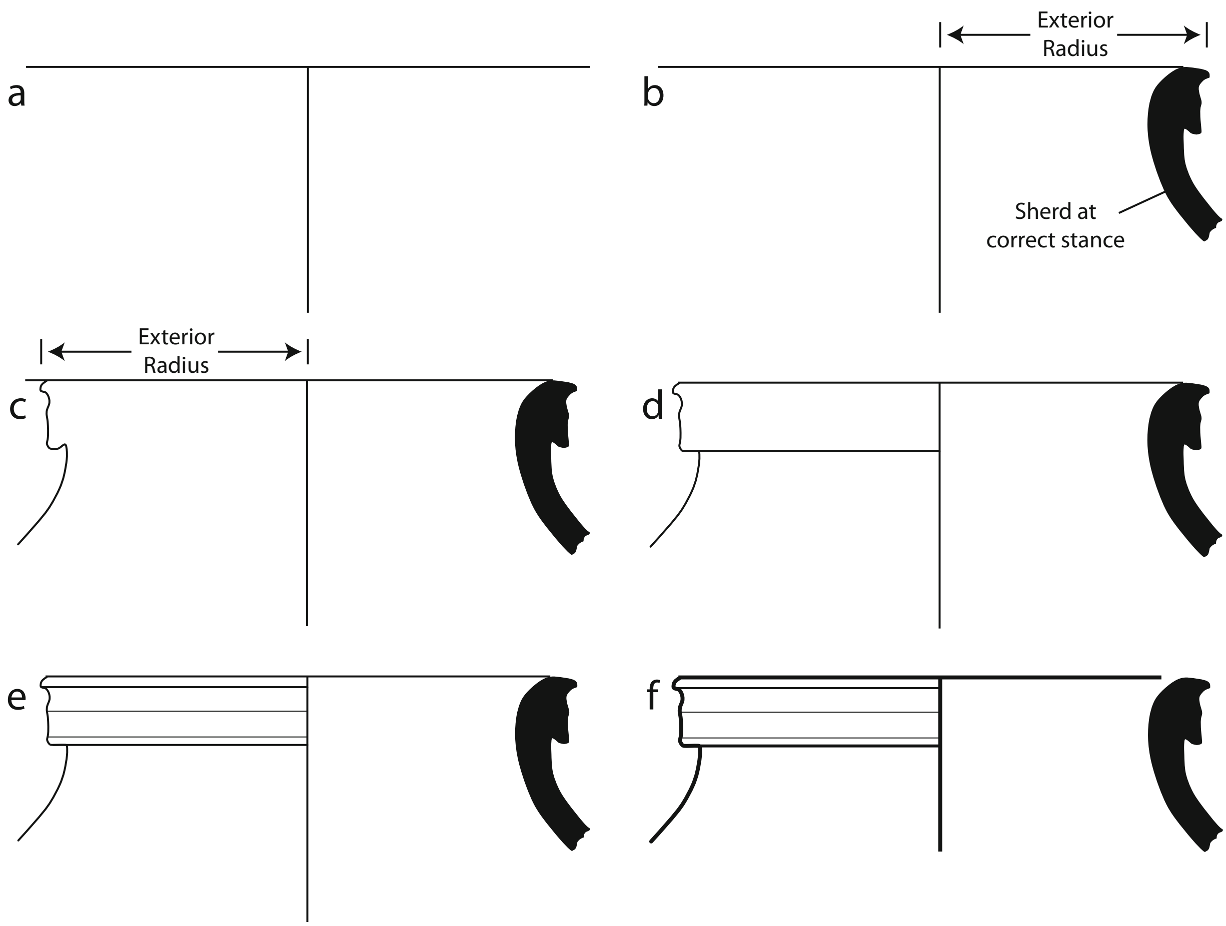


Fig. 21.14 Steps in creating a reconstructed view of a vessel, from the vertical axis and stance line (**a**), addition of the radial section (**b**) at the correct distance from the center line, reflection of the outer line of the section (**c**), correction of the view of the overhanging lip (**d**), addition of carination lines or rim details (**e**), and final view with standardized line thicknesses and gap between rim and stance line (**f**)

parts of a profile, such as a folded-over rim, it is important to depict it on this reconstruction as it would look from the outside, not in cross-section (Fig. 21.14d).

Pottery drawings executed on paper then need to be inked, by tracing the drawings onto a more stable medium such as Mylar™ with technical pens of the right point size for the project's style. Today, it is more common to "ink" them on a computer, but most of the same issues apply. Generally, the stance line, exterior profile, and line of the central axis should be thicker than the lines for carinations, grooves and small details, and the stance line should be interrupted at the end close to the cross-section. A small gap there is necessary to ensure that the stance line will not obscure the shape of the lip and, for printed illustrations, ink will not "bleed" from the stance line to the lip. The cross-section should either be filled solid (usually black) or hatched, but one should avoid "busy" fills for the section unless they have a clear purpose, such as indicating coil boundaries. For vessels that have handles, it is usual to add cross-sections of the handles in appropriate places to show whether they are strap-shaped, oval, cylindrical, or more complex in shape (Fig. 21.16). In some cases, there may also be reason to show either top or bottom views

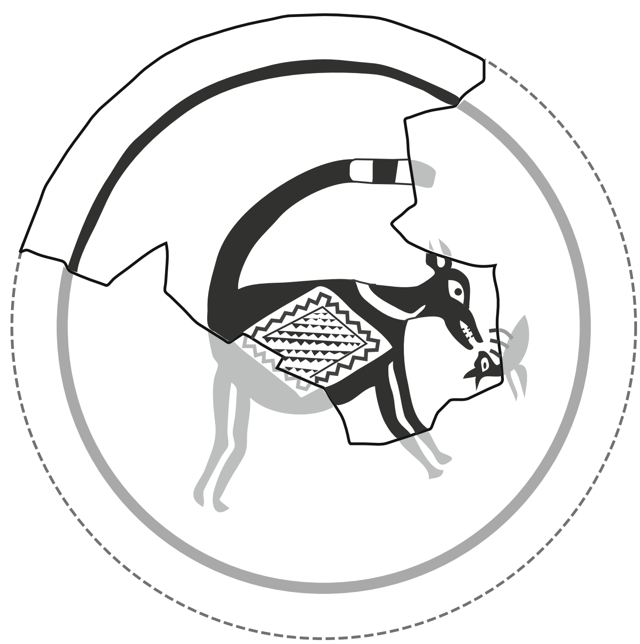


Fig. 21.15 Top (transverse) view of a decorated Mimbres bowl, with reconstructed portions of the painted design indicated as uncertain

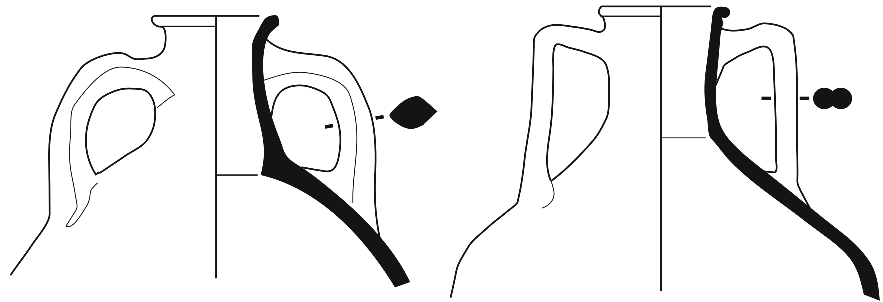


Fig. 21.16 For vessels with loop handles, it is helpful to show cross-sections to clarify their shapes, especially when they are not circular in section

to reveal significant details that do not show up from the side. Illustrating glass vessels involves exactly the same steps as ceramic ones (Fig. 21.17).

21.6 Illustrating "Small Finds"

Artifacts that are neither lithics nor vessels are so varied that no brief instructions could apply to all. It is possible to use conventions similar to those for lithics to illustrate many metal and bone tools and ornaments, and ones similar to those for vessels to illustrate some beads and other objects that have radial symmetry, but many kinds of artifacts require their own sets of conventions. For example, it is usual to depict Near Eastern cylinder seals as though they have been rolled out, to show the whole design, and indeed rolling them on modelling clay is a usual step in their illustration.

Unlike pottery and lithics, which are usually drawn at 1:1 scale and then reduced for publication, some artifacts, such as buttons and ornaments, are often small enough to call for drawing at an enlarged scale, such as 2:1 or 4:1, so that small details are clearer.

21.7 3D Scans of Artifacts

It is becoming increasingly common for 3D scanners and inexpensive photogrammetry to be available to archaeologists. We can use these to create digital models of artifacts that are themselves a record of the artifacts' forms and textures (Kelley and Wood 2018), but we can also use them as the basis for traditional, 2D illustrations (Dryer and Mazierski 2009; Gilboa et al. 2013).

After scanning an artifact, we can digitally "slice" a digital model of it in any plane we choose. For example, we can virtually slice a radial section of a pottery vessel rather than having to slice the sherd physically on a lapidary saw. It is then possible to manipulate this section in the same way as the scan of the traced sherd, as discussed above, making a reflection about the vertical axis, and even stancing the section in the software instead of physically. This can really improve the speed and accuracy of pottery illustration. Better yet, we can rely on software to automate the orientation and drawing of sherds with minimal intervention by an illustrator (e.g., Karaskik and Smilansky 2008; Wilczek et al. 2018).

Similarly, for lithics, the 3D model can be a good basis for creating the outline for ventral and dorsal views, as well as for accurate delineation of the flake scars. This reduces the need

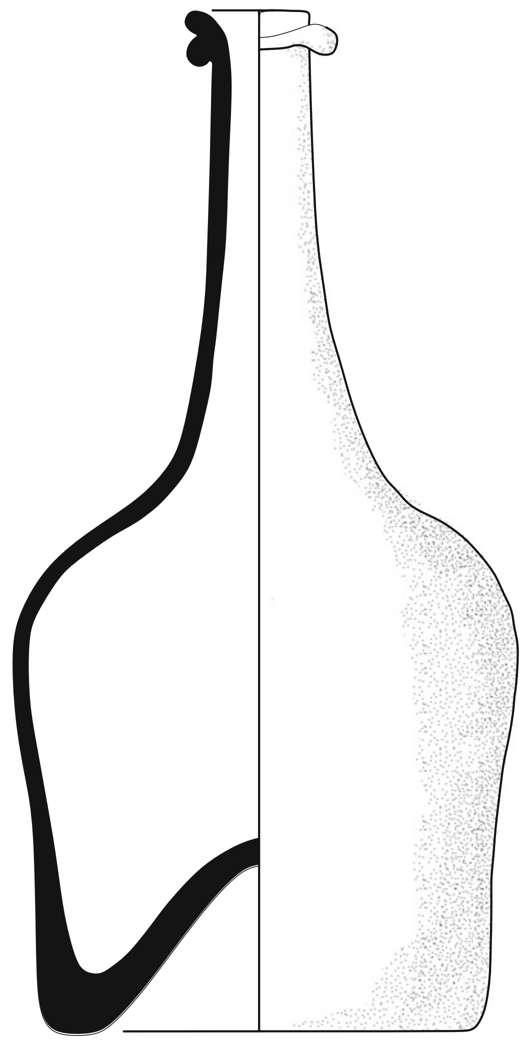


Fig. 21.17 Illustrations of glass vessels use the same conventions as for pottery vessels, here a wine bottle (Y. Salama)

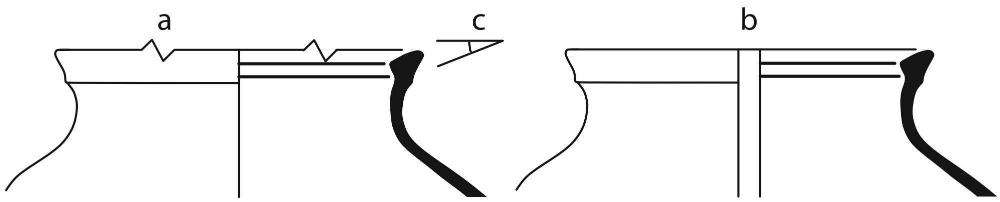


Fig. 21.18 Attempts to indicate uncertainty in a pottery illustration, showing error in the diameter by a 'zig-zag' in the stance line (**a**) or estimated error by the width of the central rectangle (**b**) and uncertainty in stance by an angle that shows an estimated error of 18° (**c**)

for a great many caliper measurements and greatly increases the speed and accuracy of the drawings (Dryer and Mazierski 2009). As with pottery, it is possible to use the 3D models themselves to generate lithic illustrations directly (Magnani 2014).

21.8 Indicating Error or Uncertainty in Illustrations

Technical illustrations of artifacts are not perfect and, as with any representation of data, their attributes are subject to error in both the measurements that led to the drawings and illustrators' attempts to represent artifacts' characteristics. Somewhat like indicating statistical errors for data points in a scatterplot with error bars, it is possible to give some indication of confidence in an artifact's representation.

For some classes of illustrations, including those of lithics, it may suffice to say that the artifacts illustrated were measured with calipers to a tolerance of ±1 mm, for example. Our confidence in illustrations that involve reconstruction is rather less than this. For pottery and other vessels, one obvious source of error is the assumption that they are circular in plan, except when clearly otherwise. Small departures from circularity, in conjunction with the somewhat large intervals on the diameter chart, likely result in errors in measurement of diameter on the order of at least ±1 cm for most vessels, and greater errors for ones that have somewhat "wonky" rims that are hard to stance. One option is just to say that the tolerance in diameter for most of the illustrations was, say, 1.5 cm, and then to indicate the lower confidence in the exceptional sherds by some convention. One possible convention for indicating uncertainty in the diameter is to put a zig-zag in the stance line (Fig. 21.18). An alternative when we have a better idea of the magnitude of error in the diameter is to replace the thick line of the center axis with a band whose width represents the estimated error. Another attribute of vessel illustrations that is prone to error is the stance angle. Rims and bases are almost never perfectly straight and only when the sherd preserves a reasonably large proportion of the circumference can we measure stance very precisely and accurately. If a rim or base is somewhat wavy, interrupted by knobs or handles, or only preserved over a short horizontal distance, stance can be very difficult to estimate. Archaeologists have yet to agree on a convention to indicate this uncertainty, but one possibility would be to place a small angle near the rim or on the stance line to indicate its likely magnitude in the most uncertain cases (Fig. 21.18).

21.9 Layout of Plates

An archaeological "plate" is typically a page that illustrates several or many individual artifacts in one place, facilitating comparison, showing the diversity of artifacts found in a particular context, or displaying the range of variation in an artifact type. As with other aspects of illustration, there are conventions for how to arrange illustrations in these plates, largely for aesthetic reasons.

First, all the artifacts on the same plate should normally be at the same scale. This is to facilitate comparison. Where one or two of the objects shown are so large as to make this impractical, the large ones should be reduced further and positioned in such a way (and perhaps marked off by a line or box) as to make it obvious that they are separate from most of the items on the plate. Of course, these exceptional artifacts need to have their own scale.

Unless there are reasons, such as intent to show stratigraphic order, to do otherwise, generally you should arrange most kinds of artifacts from the smallest ones at the top to the largest at the bottom or, for vessels, from open forms, such as cups and bowls, to closed forms, such as jars.

Wherever possible, the images should be aligned horizontally, equally spaced, and the outer edges of the outermost illustrations should be aligned to the plate margins.

21.10 Maps and Plans

No matter what styles archaeologists use for maps and plans, any map is a simplified and distorted model of reality (Monmonier 1991). Maps stretch and bend reality by fitting three-dimensional phenomena onto two-dimensional surfaces, and omit many details and add new ones that are really someone's interpretations. This is not only true of maps created the old-fashioned way, with tapes, theodolites, or total stations, but also those made with the aid of drones or photogrammetry (see also Gillings et al. 2018).

The fact that maps and plans are usually reduced for publication even more than artifact illustrations, makes it all the more important to think carefully about the degree of detail and precision that makes sense and what the line thicknesses will look like after this reduction. Some archaeologists exert considerable effort in the field (or, increasingly, in the lab) to ensure that field plans have precision of about ± 1 cm or even less, forgetting that the thickness of a pencil line at a scale of, say, 1:200 represents 20 cm. Whether or not that degree of precision will be visible on the published plan, you should also think about how meaningful it is. It is not advocating sloppiness in field measurement to point out that misplacing a few small stones on a map by 1 or 2 cm would be very unlikely to alter anyone's interpretation of archaeological context in a meaningful way. Consequently, archaeologists need to balance meaningful precision with the costs in time that could be invested elsewhere.

Everything that applies to graphs (Chap. 5), such as data: ink ratio and chartjunk (Tufte 1983), is equally applicable to maps, plans and sections. Avoid making maps too complicated, crowded or misleading. A simple map that gives prominence to the information you are trying to convey is much more effective than one that crams in so much information that its point is lost.

Every map should include several key features. One is a scale. Saying that the map scale is 1:200 or the like is pointless if there is any chance that publication will alter that scale. Instead, there should be a scale bar, and having templates for scales in various sizes makes it easy to paste them into map files and ensure that they adhere to a consistent style. The scale should be simple and have no more labelled increments than are necessary to give viewers a sense of magnitude. Most maps also require a simple, tasteful North arrow, preferably pasted from a template that once again ensures consistency in style. Among the exceptions to this rule are maps of polar regions, as the North direction varies so much over the map that it is better to use lines of longitude. The scale and North arrow should occupy parts of the map where they will not be obtrusive. For a map of a small region whose location might not be obvious, it is good to put an inset map in one corner to show where it is. Plans of excavation areas should show either labelled grid lines that, much like lines of longitude and latitude, help to locate the excavation area in space, or at least the labelled corners of the grid (marked by circles or crosses). Excavation plans should also show the locations of benchmarks that were used as reference points for mapping, or control points for photogrammetry.

Maps and plans typically require a variety of symbols or styles to indicate such things as materials (e.g., stone, clay, brick), elevations, actual and reconstructed parts of features, find-spots of individual artifacts or sampling locations, and areas of post-depositional disturbance. Some of these will require a key—a sort of table that shows what each color, hatch style, or icon symbolizes—while ones used less often might call for individual labels. Elevations are often indicated on maps by "contour lines" —curves following points of equal elevation above or below sea level—while "spot elevations" are points on the map, such as control points, high points of walls, or low points of pits and trenches, where someone measured an elevation in the field. Typically, spot elevations appear as large dots or small crosses or triangles with text in small point size next to them to indicate their height (e.g., Fig. 21.19). In plans of architectural remains, archaeologists sometimes reconstruct the probable extent of walls that have not survived or not been excavated with conventions such as dashed lines or greyed fills (Fig. 21.20).

21.11 Stratigraphic Sections and Architectural Elevations

These are a lot like maps and plans except that they show vertical, rather than horizontal planes, so most of the same guidelines apply. Again, they represent some archaeologist's interpretation, such as an excavator's interpretation of stratigraphic relationships among interfaces and deposits that intersect some vertical plane of the excavation. Consequently, they vary considerably in style (Harris 1979).

Every stratigraphic section drawing should include certain elements. Rather than illustrating sediments realistically, sections usually use sets of conventions to represent an archaeologist's or geoarchaeologist's understanding of the stratigraphy. There must be some indication of where the section fits in space. That means including at least one horizontal line labelled with its elevation above sea level or including a vertical scale (like a y-axis) labelled in meters above sea level. There should be either vertical lines to indicate the borders between units in the site grid, or a key map that shows the location of the section, typically with its two ends labelled (e.g., A and A'). The vertical and horizontal scales should be obvious, but not over-labelled, and there should be clear indication of the direction of view, such as "North Section," to help viewers orient themselves. It is usual to have a convention to indicate unexcavated sediments, as distinct from bedrock, a key to the colors, hatching or shading used to distinguish different kinds of sediments, and labels for each stratigraphic layer or unit.

For archaeological sites that have well-preserved architecture, it is often desirable not only to have plan views as described above, but also detailed drawings of how walls were constructed, "elevations" that show side-views of the

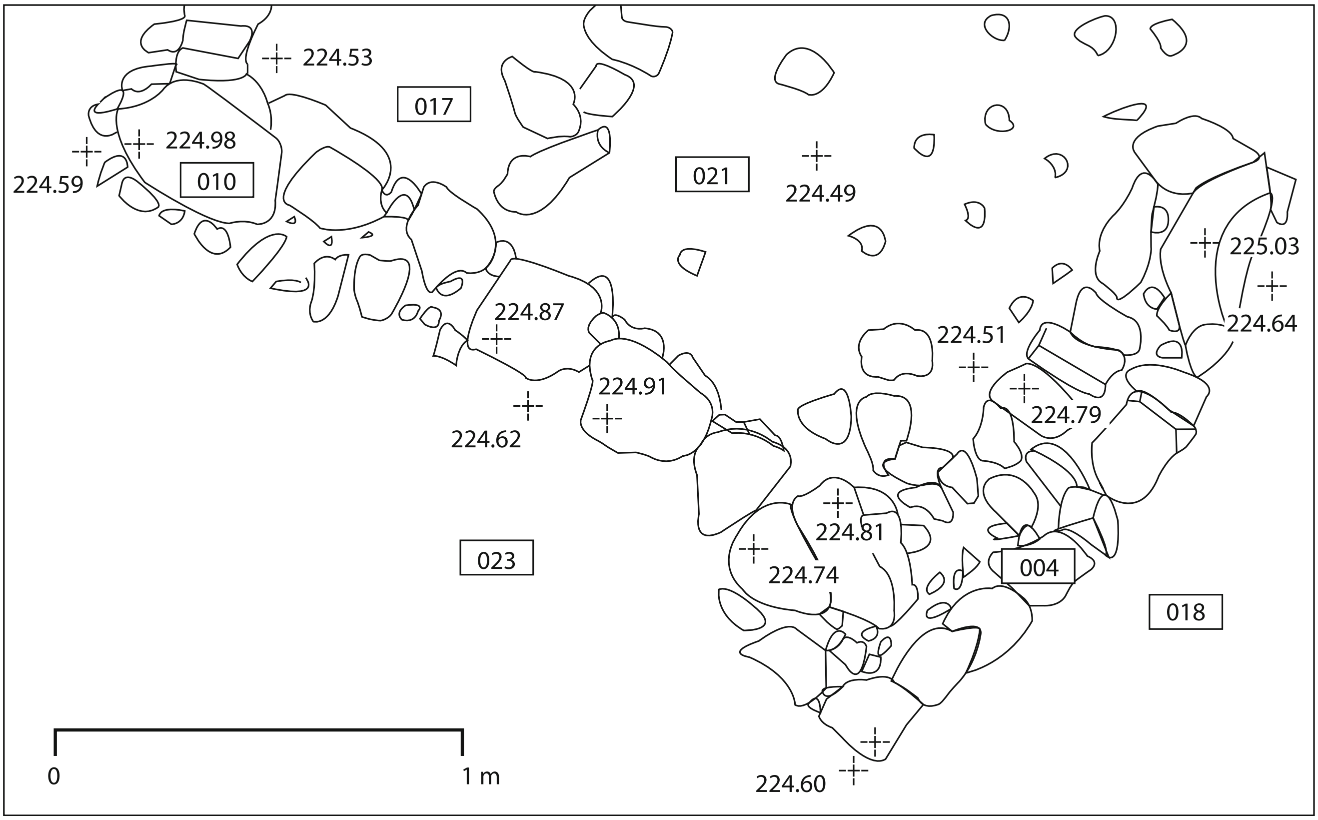


Fig. 21.19 Spot elevations on a map of an excavation area to indicate the heights of stone walls relative to floor contexts. Small numerals are heights in meters above sea level, while large numerals in boxes are context or "locus" numbers

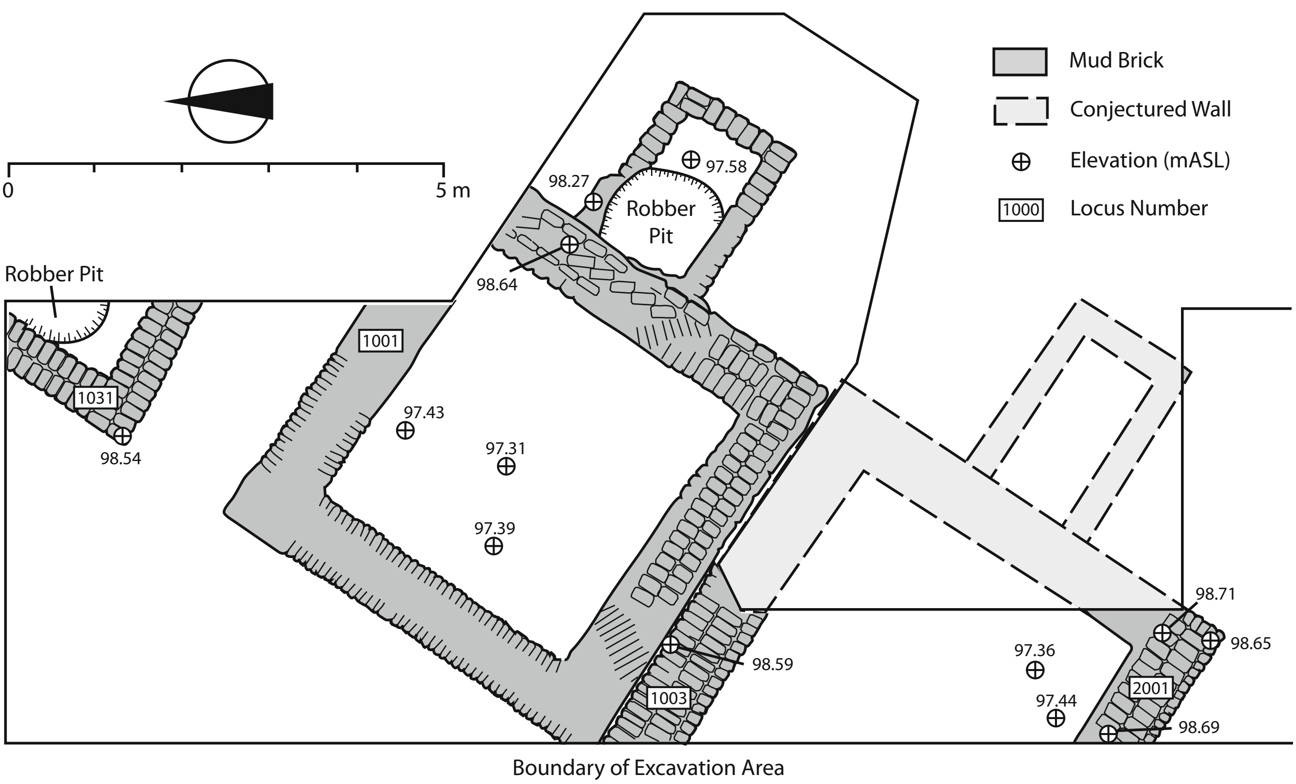


Fig. 21.20 Reconstructed or conjectural walls in a plan of mud-brick tomb structures at Tell al-Maskhuta, Egypt. (Modified from Holliday 1982: 151)

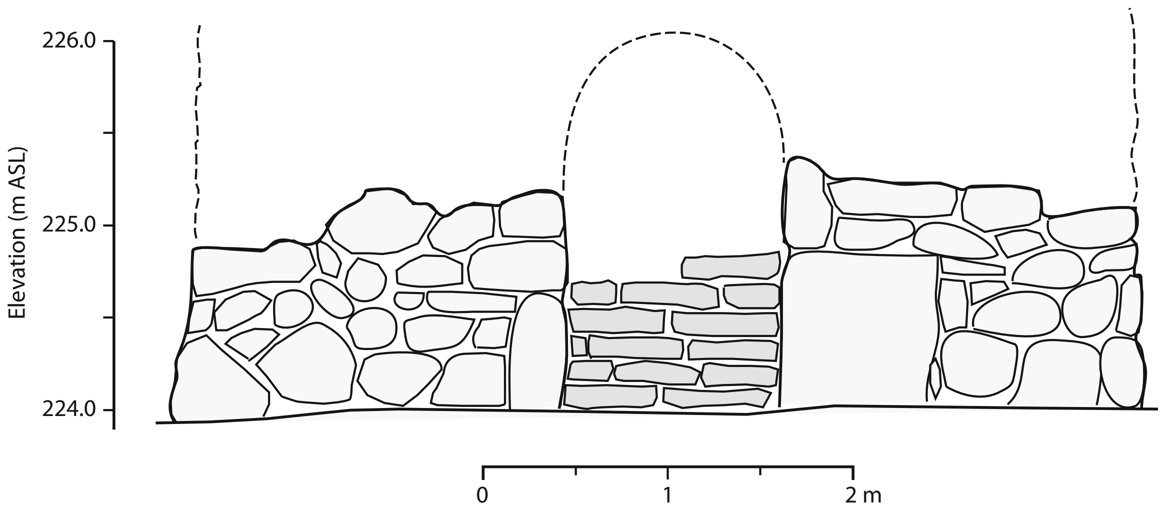


Fig. 21.21 Elevation drawing of an excavated dry-stone wall with a blocked doorway and conjectural reconstruction of missing portions

preserved portions of walls and buildings (Fig. 21.21), and sometimes illustrations that reconstruct the likely appearance of portions that are not preserved. These often follow much the same conventions that an architect would use.

21.12 Reconstructions and Visualization

Many archaeological projects depart from the arguably sterile world of technical illustration to provide views that purport to reconstruct aspects of daily life in the past. These more life-like illustrations are very important for conveying archaeologists' interpretations of how ancient material culture "worked" in its cultural context, and also how people may have interacted (Dobie 2019; Hodgson 2002).

21.13 Digital Illustration

The wide availability of graphic, illustration, GIS, and CAD (Computer-Aided Design) software has been a boon to archaeology (Steiner 2005). Whether we scan pencil drawings or use digital photographs, photogrammetry, RTI, or 3D scans to make a base drawing in a computer file, or make the image completely on a computer, these software packages provide many useful features, but also require some caution in use.

Scanned 2D images and digital photographs are bit-mapped or **raster** images, meaning that they are made up of lots of pixels that vary in color or greyscale, and tend to have very large file sizes. Generally, our goal is to use them as the basis for making **vector**-based graphics, which more efficiently represent shapes by mathematical descriptions of lines, curves, and line thicknesses, not only reducing file size, but resulting in smoother-looking images that look good at different scales. Accomplishing this involves opening the file of a bit-mapped image and then tracing it, either manually with a cursor moved with a mouse or digitizing tablet, or automatically with auto-trace routines in some of the software. However, auto-trace programs are not very discriminating, so it is often better to trace shapes manually. Rather than trace lines and curves continuously, it is often better to "click" at intervals along the shape; the software will fill the intervals with curves or line segments that can later be smoothed into curves and, after some experience, most users get pretty good at estimating how densely they need to click points to get reasonably good fit to the shapes they are tracing.

Most of these software packages allow users to make "layers" that they can turn on or off. This is really important. In tracing a bit-map, the bit-map should be left untouched, in its own layer, while tracing in one or more other layers. Turning the base layer's visibility on and off makes it easy to check on the progress of the drawing and to see more easily what portions are still untraced. Discarding the base layer when the drawing is finished reduces file size. In illustrating a stratigraphic section or making single-context maps, each layer or stratigraphic unit can have its own layer in the software, making it possible to show all or only some of them in different versions of the drawing. For all types of illustrations, things like scales and North arrows can go in one layer, labels in another layer, and things like inset maps or keys in still other layers. In anticipation of making a lot of different kinds of maps of the same area, such as maps of a site, it is very helpful to make a base map or template that has essential features, such as site grid, boundaries of excavated area, and scale, in their own layers.

This template can then serve many times as the basis for maps that show different stratigraphic phases, distributions of different artifact types, refits among lithics or pottery, or magnetometry results, to name only a few examples. This not only saves a lot of time and effort, it also ensures adherence to a consistent style. Use of the layers also makes it easier to make both color and black-and-white or greyscale versions of the same illustration. Today, many journals use color images for their online versions but not for their print versions, and you can easily accommodate this by turning the color layers on or off.

Not surprisingly, most graphics software also provides many other useful features, including ease of creating rectangles and other polygons, stippling, shading and shadows, and numerous choices of fonts, line styles and weights, arrowheads, and icons. However, "gimmick" features that do not contribute to viewers' understanding of the graphics should be avoided.

When a digital image is complete, it makes sense to delete the temporary layer or layers with the original bit-mapped image and then save to the file format required for publication (Table 21.1). File formats that ensure that finished illustrations are not distorted by rescaling, especially in their line weights, are much preferable. If in doubt, test it out by printing hard copies at different scales. As with any digital files, it is critical to take precautions to back up the images and to use file names that facilitate finding and identifying images quickly and reliably. For artifact illustrations, this usually means including the artifact number in the file name.

21.13.1 Digital File Formats

The final file format of a digital image depends on its ultimate purpose. A web site, for example, will usually demand return to a raster image, such as jpeg. For publication in a book or journal, however, the publisher will generally dictate one or a small number of acceptable formats, often including both the format of the graphic software you used to create the drawing, to allow any final editing of the image, and another format that will "freeze" the image in its current form (or fixed-document format) for use, for example, by reviewers. Table 21.1 indicates a variety of common file formats for images and how they are typically used.

21.14 The Publication Process

One of the final steps in the research process is publishing results. Increasingly, this is online in the form of web sites and online databases, but more traditional publication media are still very important.

The first step in publishing an article in a journal, still the mainstay of archaeological publication, is for the author or corresponding author (for multiple-authored work) to submit the manuscript, usually electronically through an online editorial system. Then the editor will decide whether to reject it outright (often because of poor fit to the journal's mission) or send it out for peer review. Generally, two or three reviewers with expertise in the subject matter of the paper will read the manuscript and make recommendations to reject it, ask for

Table 21.1 Common 2D file formats for archaeological illustrations. Mixed types use both vector and pixel data

Format	Type	File size	Name and purpose
.ai	Vector	Medium	Adobe Illustrator™ file, for saving illustrations in progress and finished ones. Can be edited
.bmp	Bit-map	V large	Windows™ image file, uncompressed
.bpg	Compressed raster	Small	Better Portable Graphics, a new alternative to .jpg with smaller file sizes for the same image quality
.eps	Mixed	Small	Encapsulated Postscript, basis for early versions of Adobe™ software, used for fixed-layout documents for publishing print versions of text and images
.gif	Compressed raster	Medium	Graphics Interchange Format, for simple images with few colors and for animation
GLE	Vector		Graphics Layout Engine, used for graphs, charts and mathematics, outputs to .eps, .jpg, .pdf and .png
.jpg	Compressed raster	Medium	For photographs and images on web pages. Need to tailor pixel resolution to final use
.pdf	Mixed	Medium	Portable Document Format, an Adobe™ format now in public domain that is independent of applications and operating systems, for fixed-layout, flat documents as well as fillable forms
.png	Compressed raster	Large	Portable Network Graphics, open-source alternative to .gif for images with large monocolored areas. Accommodates greyscale and good for web sites, does not lose detail during compression
.svg	Vector	Small	Scalable Vector Graphics, excellent for web sites
.tif	Compressed raster	Large	Tagged Image File, common for digital photographs but not suitable for web sites

major or minor revisions, or to accept it. Peer review is the backbone of scholarly publication and is intended to ensure the quality of the resulting publication by providing authors with constructive feedback. However, reviewers are human, not always familiar with all aspects of the subject matter, and thus not infallible. The editor reads the reviewers' recommendations and makes a decision, not necessarily following the reviews, especially when they are mixed. Usually, the editor asks for some or all of the revisions that one or more reviewers recommended. Having made these revisions or addressed reviewers' concerns (sometimes challenging a reviewer's comment), the corresponding author re-submits the manuscript, when it may go back to the same reviewers or to different ones before the editor reaches a final decision.

Once accepted, the manuscript goes to production. Editors and layout specialists will edit the text for style, contact the corresponding author with questions about unclear passages, and arrange the text and images on pages as they will appear in the final publication, usually in .pdf format. When they have finished this apart, usually, from final pagination, they send "proofs" to the corresponding author, who is responsible to check it carefully for errors and make small changes as needed, either using electronic mark-up on the file or by annotations in the text and margins of a hard copy (Table 21.2). This is not the time to ask for any major changes, especially ones that would alter the pagination. Correcting proofs is also on a tight timeline. Generally, corresponding authors should correct the proof and return it to the editor within 1 week. The final steps in production are to implement the proof corrections, add correct pagination, and send the fully formatted file for printing, posting online, or both. The process for books and monographs is similar to that for journal articles.

Table 21.2 A few examples of common marginal symbols for marking up a publisher's proof

Symbol	Meaning
∧	Insert or make subscript
	Delete
	Close up
ital	Italic
¶	Begin new paragraph
fl	Flush left
tr	Transpose letters
∧ ,	Insert comma
lc	Lower case
caps	Capital letters

21.14.1 Publishers' Styles

When you submit research for publication, both text and images must adhere to publishers' requirements. Follow them exactly, including how you format bibliography and citations and any specifications regarding illustrations or the file formats in which you should save them. *The Journal of Field Archaeology* provides particularly detailed advice on illustrations (JFA n.d.), and all archaeological journals post "guides to contributors" that outline all their requirements for abstract, key words, citation, bibliography, language, spelling, units of measurement, images, file formats, and how to submit the files to their online submission systems.

Aside from these guidelines, it is useful to predict how much the publisher is likely to reduce any graphics. It is an excellent idea to look at some recent issues of the journal or monograph series to examine the format. If it is a two-column format, for example, tall or square graphics are likely to occupy one column, while wider (landscape) ones are likely to spread across two columns. Since likely reduction scale is important for the choice of line weights and the point sizes of labels on your images, this is important for finalizing illustrations.

21.14.2 Online Publication and Unconventional Publication

Today, a good deal of archaeological publication takes place in online venues. Some of these are much like traditional, print-based journals, but with "Open Access" papers in .pdf format. However, many of these publications are web sites and we can anticipate the growing importance of "dynamic publications" (Rahtz et al. 1992).

Web sites provide venues for broad and free dissemination of potentially large amounts of text and images and, as a form of hypermedia (p. 55), allow authors to link information in ways that encourage readers to follow their own interests, rather than reading material linearly. However, they do not give those readers entirely free rein either. In other words, authors anticipate the kinds of links that readers will likely find of interest, or even the ones they would like to encourage readers to explore. Some of these links can be to videos or simulations, rather than text or still images. It is also possible to give readers access to the content of a relational database via a web site. This last option allows users to explore the data in even more diverse ways.

21.15 Summary

- Archaeological illustrations are not just pictures, they involve a graphical information language that omits some details, while highlighting or reconstructing others in a standardized way, to describe and interpret artifacts or their contexts
- Photographs and even 3D models are not sufficient, alone, to portray this information
- Each project should formulate "styles" for its illustrations of artifacts, maps, plans and sections to ensure consistency both aesthetically and in terms of information content
- Styles include specific guidelines on line weights, fonts, point sizes, fills or hatching, North arrows, scales and other details that will appear regularly in project graphics
- Most artifact illustrations should use orthographic projection for multiple views, shown as though lit from upper left
- Pottery and other vessels should be shown as though reconstructed, with a cutaway view so that the interior, exterior, radial section and vessel diameter are all visible
- Some aspects of illustration typically involve "old-fashioned" drawing and measuring instruments, but also digital imaging
- Illustrations are just one aspect of the publication process, which often includes peer review and proof correction
- Dissemination of results occurs in conventional printed monographs and journals, but also in online journals, web sites, and online databases

References Cited

Addington, L. R. (1986). *Lithic illustration: Drawing flaked stone artifacts for publication*. Chicago: University of Chicago Press.

Adkins, L., & Adkins, R. A. (1989). *Archaeological illustration*. Cambridge: Cambridge University Press.

Cambden, W. (1806). *Chorographical description of the flourishing kingdoms of England, Scotland and Ireland and the islands adjacent from the earliest antiquity* (Vol. 1, 2nd ed.). London: John Stockdale.

Chase, P. G. (1985). Illustrating lithic artifacts: Information for scientific illustration. *Lithic Technology, 14*(2), 57–70.

Dillon, B. (1985). *Student's guide to archaeological illustrating* (2nd ed.). Los Angeles: Cotsen Institute of Archaeology.

Dobie, J. (2019). *Illustrating the past*. Swindon: Historic England.

Dryer, M., & Mazierski, D. (2009). Illustrating artifacts using digital rendering techniques. *JBC, 35*(2), E35–E41.

Gilboa, A., Tal, A., Shimshoni, I., & Kolomenkin, M. (2013). Computer-based, automatic recording and illustration of complex archaeological artifacts. *Journal of Archaeological Science, 40*, 1329–1339.

Gillings, M., Hacıgüzeller, P., & Lock, G. (Eds.). (2018). *Re-mapping archaeology: Critical perspectives, alternative mappings*. London: Routledge.

Graser, A., & Peterson, G. N. (2018). *QGIS map design* (2nd ed.). Chugiak: Locate Press.

Griffiths, N., Jenner, A., & Wilson, C. (2007). *Drawing archaeological finds: A handbook* (Revised ed.). London: Archetype Publications.

Harris, E. C. (1979). *Principles of archaeological stratigraphy*. London: Academic.

Hodgson, J. (2002). *Archaeological reconstruction: Illustrating the past*. Reading: Institute of Field Archaeologists.

Holliday, J. S. (1976). A technical aid to pottery drawing. *Antiquity, 50*, 223–229.

Holliday, J. S. (1982). *Cities of the Delta III, Tell el-Maskhuta* (American Research Center in Egypt Reports 6). Malibu: Undena Publications.

James, S. (2015). 'Visual competence' in archaeology: A problem hiding in plain sight. *Antiquity, 89*(347), 1189–1202.

JFA (*Journal of Field Archaeology*) (n.d.) Instructions for authors. https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=yjfa20. Accessed 28 Dec 2018.

Karaskik, A., & Smilansky, U. (2008). 3D scanning technology as a standard archaeological tool for pottery analysis: Practice and theory. *Journal of Archaeological Science, 35*, 1148–1168.

Kelley, K., & Wood, R. K. (Eds.). (2018). *Digital imaging of artifacts: Developments in methods and aims*. Oxford: Archaeopress.

Lopes, D. M. (2018). Aesthetics on the edge: Where philosophy meets the human sciences. Oxford Scholarship Online. Oxford: Oxford University.

Lü Dalin (1092). *Kaogutu* (Illustrated Investigations of Antiquity).

Magnani, M. (2014). Three-dimensional alternatives to lithic illustration. *Advances in Archaeological Practice, 2*(4), 285–297.

Monmonier, M. (1991). *How to lie with maps*. Chicago: University of Chicago Press.

Newlands, D. L., & Breede, C. (1976). *An introduction to Canadian archaeology*. Toronto: McGraw-Hill Ryerson.

Raczynski-Henk, Y. (2017). *Drawing lithic artefacts*. Leiden: Sidestone Press.

Rahtz, S. P. Q., Hall, W., & Allen, T. (1992). The development of dynamic archaeological publications. In P. Reilly & S. Rahtz (Eds.), *Archaeology and the information age: A global perspective* (pp. 360–383). London: Routledge.

Shirvaklar, P. (2017). Analytical drawing. In A. M. W. Hunt (Ed.), *The Oxford handbook of archaeological ceramic analysis* (pp. 217–230). Oxford: Oxford University Press.

Steiner, M. (2005). *Approaches to archaeological illustration: A handbook*. London: Council for British Archaeology.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire: Graphics Press.

Wilczek, J., Monna, F., Jebrane, A., Chazal, C. L., Navarro, N., Couette, S., et al. (2018). Computer-assisted orientation and drawing of archaeological pottery. *Journal of Computing and Cultural Heritage, 11*(4), 1–17.

Index

E. B. Banning, *The Archaeologist's Laboratory*, Interdisciplinary Contributions to Archaeology,
https://doi.org/10.1007/978-3-030-47992-3