Part I

Building Blocks

Chapter One

Preliminaries

Math is a formal language useful in clarifying and exploring connections between concepts. Like any language, it has a syntax that must be understood before its meaning can be parsed. We discuss the building blocks of this syntax in this chapter. The first is the variables that translate concepts into mathematics, and we begin here. Next we cover groupings of these variables into sets, and then operators on both variables and sets. Most data in political science are ordered, and relations, the topic of our fourth section, provide this ordering. In the fifth section we discuss the level of measurement of variables, which will aid us in conceptual precision. In the sixth section we offer an array of notation that will prove useful throughout the book; the reader may want to bookmark this section for easy return. Finally, the seventh section discusses methods of proof, through which we learn new things about our language of mathematics. This section is the most difficult, is useful primarily to those doing formal theory or devising new methods in statistics, and can be put aside for later reading or skipped entirely.

1.1 VARIABLES AND CONSTANTS

Political scientists are interested in concepts such as participation, voting, democracy, party discipline, alliance commitment, war, etc. If scholars are to communicate meaningfully, they must be able to understand what one another is arguing. In other words, they must be specific about their theories and their empirical evaluation of the hypotheses implied by their theories.

A theory is a set of statements that involve concepts. The statements comprise assumptions, propositions, corollaries, and hypotheses. Typically, assumptions are asserted, propositions and corollaries are deduced from these assumptions, and hypotheses are derived from these deductions and then empirically challenged.[1] Concepts are inventions that human beings create to help them understand the world. They can generally take different values: high or low, present or absent, none or few or many, etc.

Throughout the book we use the term "concept," not "variable," when discussing theory. Theories (and the hypotheses they imply) concern relationships among abstract concepts. Variables are the indicators we develop to measure

[1] Of course, assumptions and the solution concepts from which deductions are made may be empirically challenged as well, but this practice is rarer in the discipline.

our concepts. Current practice in political science does not always honor this distinction, but it can be helpful, particularly when first developing theory, to speak of concepts when referring to theories and hypotheses, and reserve the term variables for discussion of indicators or measures.

We assign *variables* and *constants* to concepts so that we may use them in formal mathematical expressions. Both variables and constants are frequently represented by an upper- or lowercase letter. Y or y is often used to represent a concept that one wishes to explain, and X or x is often used to represent a concept that causes Y to take different values (i.e., vary). The letter one chooses to represent a concept is arbitrary—one could choose m or z or h, etc. There are some conventions, such as the one about x and y, but there are no hard-and-fast rules here.

Variables and constants can be anything one believes to be important to one's theory. For example, y could represent voter turnout and x the level of education. They differ only in the degree to which they vary across some set of cases. For example, students of electoral politics are interested in the gender gap in participation and/or party identification. Gender is a variable in the US electorate because its value varies across individuals who are typically identified as male or female.[2] In a study of voting patterns among US Supreme Court justices between 1850 and 1950, however, gender is a constant (all the justices were male).

More formally, a **constant** is a concept or a measure[3] that has a single value for a given set. We define sets shortly, but the sets that interest political scientists tend to be *the characteristics of* individuals (e.g., eligible voters), collectives (e.g., legislatures), and countries. So if the values for a given concept (or its measure) do not vary across the individuals, collectives, or countries, etc., to which it applies, then the value is a constant. A **variable** is a concept or a measure that takes different values in a given set. Coefficients on variables (i.e., the parameters that multiply the variables) are usually constants.

1.1.1 Why Should I Care?

Concepts and their relationships are the stuff of science, and there is nothing more fundamental for a political scientist than an ability to be precise in concept formation and the statement of expected relationships. Thinking abstractly in terms of constants and variables is a first step in developing clear theories and testable hypotheses.

[2]Definitions of concepts are, quite properly, contested in all areas of academia, and gender is no exception. Though it is not a debate that generates a great deal of interest among students of participation or party identification, it will be rather easy for you to find literature in other fields debating the value of defining gender as a binary variable.

[3]By measure we mean an operational indicator of a concept. For example, the concept gender might be measured with a survey question. The survey data provide a measure of the concept.

Table 1.1: Common Sets

Notation	Meaning
$\mathbb{N}$	Natural numbers
$\mathbb{Z}$	Integers
$\mathbb{Q}$	Rational numbers
$\mathbb{R}$	Real (rational and irrational) numbers
$\mathbb{C}$	Complex numbers
Subscript: $\mathbb{N}_+$	Positive (negative) values of the set
Superscript: $\mathbb{N}^d$	Dimensionality (number of dimensions)

1.2 SETS

This leads us naturally into a discussion of sets. For our purposes,[4] a *set* is just a collection of elements. One can think of them as groups whose members have something in common that is important to the person who has grouped them together. The most common sets we utilize are those that contain all possible values of a variable. You undoubtedly have seen these types of sets before, as all numbers belong to them. For example, the counting numbers $(0, 1, 2, \ldots,$ where $\ldots$ signifies that this progression goes on indefinitely) belong to the set of natural numbers.[5] The set of all natural numbers is denoted $\mathbb{N}$, and any variable n that is a natural number must come from this set. If we add negative numbers to the set of natural numbers, i.e., $\ldots, -3, -2, -1$, then we get the set of all integers, denoted $\mathbb{Z}$. All numbers that can be expressed as a ratio of two integers are called rational numbers, and the set of these is denoted $\mathbb{Q}$. This set is larger than the set of integers (though both are infinite!) but is still missing some important irrational numbers such as π and e. The set of all rational and irrational numbers together is known as the real numbers and is denoted $\mathbb{R}$.[6]

Political scientists are interested in general relationships among concepts. Sets prove fundamental to this in two ways. We have already discussed the association between concepts and variables. As the values of each variable, and so of each concept, are drawn from a set, each such set demarcates the range of possible values a variable can take. Some variables in political science have ranges of values equal to all possible numbers of a particular type, typically either integers, for a variable such as net migration, or real numbers, for a

[4]These purposes, you will recall, are to build intuition rather than to be exact. We play somewhat loosely with ordered sets in what follows, and ignore things like Russell's paradox.

[5]Some define the natural numbers without the zero. We are not precise enough in this book to make this distinction important.

[6]You may have occasion to use complex numbers, denoted $\mathbb{C}$. These have two components, a real and an imaginary part, and can be written $a + bi$, where a and b are both real numbers and $i = \sqrt{-1}$. These are beyond the scope of this book, though amply covered by classes in complex analysis.

variable such as GDP. More typically, variables draw their values from some subset of possible numbers, and we say the variable x is an element of a subset of $\mathbb{R}$. For example, population is typically an element of $\mathbb{Z}_+$, the set of all positive integers, which is a subset of all integers. (A $+$ subscript typically signifies positive numbers, and a $-$ negative.) The size and qualities of the subset can be informative. We saw this earlier for the gender variable: depending on the empirical setting, the sets of all possible values were either $\{\text{Male}, \text{Female}\}$ or $\{\text{Male}\}$.[7] The type of set from which a variable's values are drawn can also guide our theorizing. Researchers who develop a formal model, game theoretic or otherwise, must explicitly note the range of their variables, and we can use set notation to describe whether they are discrete or continuous variables, for example. A variable is discrete if each one of its possible values can be associated with a single integer. We might assign a 1 for a female and 2 for male, for instance. Continuous variables are those whose values cannot each be assigned a single integer.[8] We typically assume that continuous variables are drawn from a subset of the real numbers, though this is not necessary.

A *solution set* is the set of all solutions to some equation, and may be discrete or continuous. For example, the set of solutions to the equation $x^2 - 5x + 6 = 0$ is $\{2, 3\}$, a discrete set. We term a *sample space* a set that contains all of the values that a variable can take in the context of statistical inference. When discussing individuals' actions in game theory, we instead use the term *strategy space* for the same concept. For example, if a player in a one-shot game[9] can either (C)ooperate with a partner for some joint goal or (D)efect to achieve personal goals, then the strategy space for that player is $\{C, D\}$. This will make sense in context, as you study game theory.

Note that each of these is termed a *space* rather than a set. This is not a typo; spaces are usually sets with some structure. For our purposes the most common structure we will encounter is a metric—a measure of distance between the elements of the set. Sets like $\mathbb{Z}$ and $\mathbb{R}$ have natural metrics. These examples of sets form one-dimensional spaces: the elements in them differ along a single axis. Sets may also contain multidimensional elements. For example, a set might contain a number of points in three-dimensional space. In this case, each element can be written (x, y, z), and the set from which these elements are drawn is written $\mathbb{R}^3$. More generally, the superscript indicates the dimensionality of the space. We will frequently use the d-dimensional space $\mathbb{R}^d$ in this book. When $d = 3$, this is called Euclidean space. Another common multidimensional element is an ordered pair, written (a, b). Unlike elements of $\mathbb{R}^3$, in which each

[7]As explained below, curly brackets indicate that the set is discrete. Continuous sets are demarcated by parentheses and square brackets.

[8]Formally, a discrete variable draws values from a countable set, while a continuous variable draws from an uncountable set. We define countability shortly.

[9]A one-shot game is one that is played only once, rather than repeatedly. You will encounter unfamiliar terms in the reading you do in graduate school. It is important to get in the habit of referencing a good dictionary (online or printed) and looking up terms. A search on a site like Google is often a useful way to find definitions of terms that are not found in dictionaries.

of x, y, and z is a real number, each member of an ordered pair may be quite different. For example, an ordered pair might be (orange, lunch), indicating that one often eats an orange at lunch. Ordered pairs, or more generally ordered n-tuples, which are ordered pairs with n elements, are often formed via Cartesian products. We describe these in the next section, but they function along the lines of "take one element from the set of all fruit and connect it to the set of all meals."

Political scientists also think about sets informally (i.e., nonmathematically) on a regular basis. We may take as an example the article by Sniderman, Hagendoorn, and Prior (2004). The authors were interested in the source of the majority public's opposition to immigrant minorities and studied survey data to evaluate several hypotheses. The objects they studied were individual people, and each variable over which they collected data can be represented as a set. For example, they developed measures of people's perceptions of threat with respect to "individual safety," "individual economic well-being," "collective safety," and "collective economic well-being." They surveyed 2,007 people, and thus had four sets, each of which contained 2,007 elements: each individual's value for each measure.[10] In this formulation sets contain not the possible values a variable might take, but rather the realized values that many variables do take, where each variable is one person's perception of one threat. Thus, sets here provide us with a formal way to think about membership in categories or groups.

Given the importance of both ways of thinking about sets, we will take some time now to discuss their properties. A set can be finite or infinite, countable or uncountable, bounded or unbounded. All these terms mean what we would expect them to mean. The number of elements in a **finite set** is finite; that is, there are only so many elements in the set, and no more. In contrast, there is no limit to the number of elements in an **infinite set**. For example, the set $\mathbb{Z}$ is infinite, but the subset containing all integers from one to ten is finite. A **countable set** is one whose elements can be counted, i.e., each one can be associated with a natural number (or an integer). An **uncountable set** does not have this property. Both $\mathbb{Z}$ and the set of numbers from one to ten are countable, whereas the set of all real numbers between zero and one is not. A **bounded set** has finite size (but may have infinite elements), while an **unbounded set** does not. Intuitively, a bounded set can be encased in some finite shape (usually a ball), whereas an unbounded set cannot. We say a set has a lower bound if there is a number, u, such that every element in the set is no smaller than it, and an upper bound if there is a number, v, such that every element in the set is no bigger than it. These bounds need not be in the set themselves, and there may be many of them. The greatest lower bound is the largest such lower bound, and the least upper bound is the smallest such upper bound.

Sets contain **elements**, so we need some way to indicate that a given element

[10] One could also view this as four sets of ordered pairs, with each pair containing a variable name and a person's perceptions, or one set of ordered 5-tuples, each with a person's name and her responses to each question, in order.

is a member of a particular set. A "funky E" serves this purpose: $x \in A$ states that "x is an element of the set A" or "x is in A." You will find this symbol used when the author restricts the values of a variable to a specific range: $x \in \{1, 2, 3\}$ or $x \in [0, 1]$. This means that x can take the value 1, 2, or 3 or x can be any real number from 0 to 1, inclusive. It is also convenient to use this notation to identify the range of, say, a dichotomous dependent variable in a statistical analysis: $y \in \{0, 1\}$. This means that *y either can take a value of 0 or a value of 1.* So the "funky E" is an important symbol with which to become familiar. Conversely, when something is not in a set, we use the symbol $\notin$, as in $x \notin A$. This means that, for the examples in the previous paragraph, x does not take the values 1, 2, or 3 or is not between 0 and 1. As you may have guessed from our usage, curly brackets like $\{\}$ are used to denote discrete sets, e.g., $\{A, B, C\}$. Continuous sets use square brackets or parentheses depending on whether they are closed or open (terms we define in Chapter 4), e.g., $[0, 1]$ or $(0, 1)$, which are the sets of all real numbers between 0 and 1, inclusive and exclusive, respectively.

Much as sets contain elements, they also can contain, and be contained by, other sets. The expression $A \subset B$ (read "A is a **proper subset** of B") implies that set B contains all the elements in A, plus at least one more. More formally, $A \subset B$ if all x that are elements in A are also elements in B (i.e., if $x \in A$, then $x \in B$). $A \subseteq B$ (read "A is a **subset** of B"), in contrast, allows A and B to be the same. We say that A is a proper subset of B in the first case but not in the second. So the set of voters is a subset of the set of eligible voters, and is most likely a proper subset, since we rarely experience full turnout. We also occasionally say that a set that contains another set is a superset of the smaller one, but this terminology is less common. The **cardinality of a set** is the number of elements in that set. Note that proper subsets have smaller cardinalities than their supersets, finite sets have finite cardinalities, and infinite sets have infinite cardinalities.

A **singleton** is a set with only one element and so a cardinality of one. The power set of A is the set of all subsets of A, and has a cardinality of $2^{|A|}$, where $|A|$ is the cardinality of A. Power sets come up reasonably often in political science by virtue of our attention to bargaining and coalition formation. When one considers all possible coalitions or alliances, one is really considering all possible subsets of the overall set of individuals or nations. Power sets of infinite sets are always uncountable, but are not usually seen in political science applications. The **empty set** (or **null set**) is the set with nothing in it and is written $\emptyset$. The **universal set** is the set that contains all elements. This latter concept is particularly common in probability.

Finally, sets can be ordered or unordered. The **ordered set** $\{a, b, c\}$ differs from $\{c, a, b\}$, but the **unordered set** $\{a, b, c\}$ is the same as $\{c, a, b\}$. That is, when sets are ordered, the order of the elements is important. Political scientists primarily work with ordered sets. For example, all *datasets* are ordered sets. Consider again the study by Sniderman et al. (2004). We sketched four of the sets they used in their study; the order in which the elements of those sets is maintained is critically important. That is, the first element in each set must refer to the first person who was surveyed, the second element must refer to the second person, and the 1,232nd element must refer to the 1,232nd person surveyed, etc. All data analyses use ordered sets. Similarly, all equilibrium strategy sets in game theory are ordered according to player. However, this does not mean all sets used in political science are ordered. For example, the set of all strategies one might play may or may not be ordered.

1.2.1 Why Should I Care?

Sets are useful to political scientists for two reasons: (1) one needs to understand sets before one can understand relations and functions (covered in this chapter and Chapter 3), and (2) sets are used widely in formal theory and in the presentation of some areas of statistics (e.g., probability theory is often developed using set theory). They provide us with a more specific method for doing the type of categorization that political scientists are always doing. They also provide us with a conceptual tool that is useful for developing other important ideas. So a basic familiarity with sets is important for further study.

For example, game theory is concerned with determining what two or more actors should choose to do, given their goals (expressed via their utility) and their beliefs about the likelihood of different outcomes given the choices they might make and their beliefs about the expected behavior of the other actor(s). Sets play a central role in game theory. The choices available to each actor form a set. The best responses of an actor to another actor's behavior form a set. All possible states of the world form a set. And so on.

Those of you who are unfamiliar with game theory will find this brief discussion less than illuminating, but do not be concerned. Our point is not to explain sets of actions, best response sets, or information sets—each is covered in game theory courses and texts—but rather to underscore why it is important to have a functional grasp of elementary set theory if one wants to study formal models. Finally, we note that Riker's (1962) celebrated game theoretic model of political coalition formation makes extensive use of set theory to develop what he calls the size principle (see Appendix I, pp. 247–78, of his book). That is, of course, but one of scores of examples we might have selected.[11]

1.3 OPERATORS

We now have formalizations of concepts (variables) and ways to order and group these variables (sets), but as yet nothing to do with them. Operators, the topic of this section, are active mathematical constructs that, as their name implies, operate on sets and elements of sets. Some operators on variables have been familiar since early childhood: addition ($+$), subtraction ($-$), multiplication ($*$ or $\times$ or $\cdot$ or just placing two variables adjacent to each other as in xy),

[11]Readers interested in surveys of formal models in political science that are targeted at students might find Shepsle and Bonchek (1997) and Gelbach (2013) useful.

and division ($\div$ or $/$). We assume you know how to perform these operations. Exponentiation, or raising x to the power a (x^a), is likely also familiar, as is taking an nth root ($\sqrt[n]{x}$), and perhaps finding a factorial (!) as well.

Other useful basic operators include summation ($\sum_i x_i$), which dictates that all the x_i indexed by i should be added, and product ($\prod_i x_i$), which dictates that all the x_i be multiplied. These operators are common in empirical work, where each i corresponds to a data point (or observation). Here are a couple of examples:

$$\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3,$$

and

$$\prod_{i=1}^{3} x_i = x_1 \times x_2 \times x_3.$$

Because they are just shorthand ways of writing multiple sums or products, each of these operators obeys all the rules of addition and multiplication that we lay out in the next chapter. So, for example, $\sum_{i=1}^{n} x_i^2$ does not generally equal $\left(\sum_{i=1}^{n} x_i\right)^2$ for the same reason that $(2^2+3^2) = 13$ does not equal $(2+3)^2 = 25$.[12] Other operators and their properties will be introduced as needed throughout the book. We present a collection of notation below in section 1.6 of this chapter.

You may be less familiar with operators on sets, though they are no less fundamental. We consider six here: *differences*, *complements*, *intersections*, *unions*, *partitions*, and *Cartesian products*. The **difference** between two sets A and B, denoted $A \backslash B$ (read "A difference B"), is the set containing all the elements of A that are not also in B: $x \in A \backslash B$ if $x \in A$ but $x \notin B$. This set comes up a great deal in game theory when one is trying to exclude individual players or strategies from consideration. The **complement** of a set, denoted A' or A^c, is the set that contains the elements that are not contained in A: $x \in A^c$ if x is not an element of A.[13] Continuing the example from above, the complement of the set of registered voters is the set of all people who are not registered voters.

Venn diagrams can be used to depict set relationships. Figure 1.1 illustrates the concepts of set difference and set complement. The shaded part of the left diagram is the set Registered Voters $\backslash$ Registered Democrats, which is read "Registered Voters difference Registered Democrats." Or, in other words, all registered voters who are *not* registered Democrats. The shaded part of the right diagram illustrates the set Registered Voters^c, which is "the complement

[12] Summations and products can also be repeated; this is known as a double (or triple, etc.) summation or product. If x_{ij} is indexed by i and j, then we could write $\sum_i \sum_j x_{ij}$ or $\prod_i \prod_j x_{ij}$. Multiple summations may be useful, for example, when employing discrete distributions in more than one dimension, or when considering more than one random variable in game theory.

[13] One can also think of the complement of a set A as the difference between the universal set and A.

of Registered Voters." Or, in other words, people who are not registered voters, since the universal set in this case is the set of All People. Both diagrams illustrate the concept of a subset: the set Registered Voters is a (proper) subset of the set All People, and the set Registered Democrats is a (proper) subset of the set Registered Voters. And both diagrams illustrate another concept: the sets Registered Voters and Registered Voters^c are collectively exhaustive, in that together they constitute the set All People, which is the universal set in this case. In general, a group of sets is **collectively exhaustive** if together the sets constitute the universal set.[14]

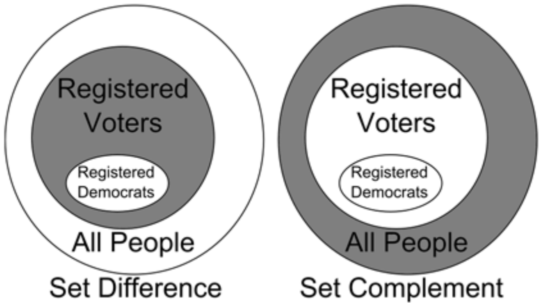

Figure 1.1: Set Difference and Complement

The **intersection** of two sets A and B, denoted $A \cap B$ (read "A intersection B"), is the set of elements common to both sets. In other words, $x \in A \cap B$ if $x \in A$ and $x \in B$. Thus, if set A consists of elected Democrats in the state of Florida and set B consists of legislators in the Florida House of Representatives, then the intersection of A and B is the set containing all Democratic House members in Florida.

The **union** of two sets is written $A \cup B$ (read "A union B") and is the set of all elements contained in either set. In other words, $x \in A \cup B$ if $x \in A$ or $x \in B$. Note that any x in both sets is also in their union. Continuing the example from above, the union of A and B is the set composed of all elected Democrats in Florida *and* all House members in Florida. Figure 1.2 shows the intersection of the sets House Members and Elected Democrats in the shaded part on the left, and their union in the shaded part on the right. The diagram on the left also illustrates the concept of mutually exclusive sets. **Mutually exclusive** sets are sets with an intersection equal to the empty set, i.e., sets with no elements in their intersection. In the diagram on the left, the two unshaded portions of the sets House Members and Elected Democrats are mutually exclusive sets. In fact, any two sets are mutually exclusive once their intersection has been removed, since they then must have an intersection that is empty.

A **partition** is a bit more complex: it is the collection of subsets whose union forms the set. The more elements a set has, the greater the number of partitions

[14]Strictly speaking, their union must equal the universal set. We discuss unions next.

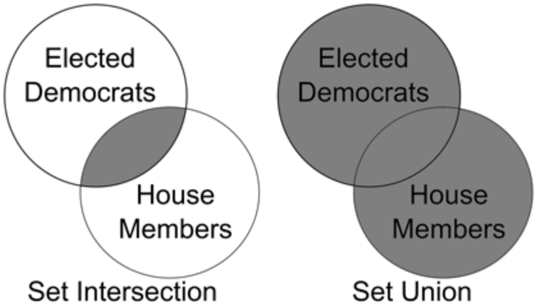

Figure 1.2: Set Intersection and Union

one can create. Let's consider the following example, the set of candidates for the 2004 US presidential election who received national press coverage:[15] $A = \{\text{Bush}, \text{Kerry}, \text{Nader}\}$. We can partition A into three subsets: $\{\text{Bush}\}$, $\{\text{Kerry}\}$, $\{\text{Nader}\}$; or we can partition it into two subsets: $\{\text{Bush}, \text{Nader}\}$, $\{\text{Kerry}\}$; or $\{\text{Kerry}, \text{Nader}\}$, $\{\text{Bush}\}$; or $\{\text{Bush}, \text{Kerry}\}$, $\{\text{Nader}\}$. Finally, the set itself is a partition: $\{\text{Bush}, \text{Kerry}, \text{Nader}\}$.

A **Cartesian product** is more complex still. Consider two sets A and B, and let $a \in A$ and $b \in B$. Then the Cartesian product $A \times B$ is the set consisting of all possible ordered pairs (a, b), where $a \in A$ and $b \in B$. For example, if $A = \{$Female, Male$\}$ and $B = \{$Income over \$50k, Income under \$50k$\}$, then the Cartesian product is the set of cardinality four consisting of all possible ordered pairs: $A \times B = \{$(Female, Income over \$50k), (Female, Income under \$50k), (Male, Income over \$50k), (Male, Income under \$50k)$\}$. Note that the type of element (ordered pairs) in the product is different from the elements of the constituent sets. Cartesian products are commonly used to form larger spaces from smaller constituents, and appear commonly in both statistics and game theory. We can extend the concept of ordered pairs to ordered n-tuples in this manner, and each element in the n-tuple represents a *dimension*. So x is one-dimensional, (x, y) is two-dimensional, (x, y, z) is three-dimensional, and so on. Common examples of such usage would be $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, which is three-dimensional space, and $S = S_1 \times S_2 \times \ldots \times S_n$, which is a strategy space formed from the individual strategy spaces of each of the n players in a game.

1.3.1 Why Should I Care?

Operators on variables are essential; without them we could not even add two numbers. Operators on sets are equally essential, as they allow us to manipulate sets and form spaces that better capture our theories, including complex inter-

[15] In August 2004 Project Vote Smart listed over ninety candidates for president of the United States, but working with the full set would be unwieldy, so we restrict attention to the subset of candidates who received national press coverage (`http://www.vote-smart.org/election_president_party.php?party_name=All`).

actions. They are also necessary for properly specifying functions of all sorts, as we shall see in Chapter 3.

1.4 RELATIONS

Now we have variables, conceptually informed groups of variables, and ways to manipulate them via operators, but we still lack ways to compare concepts and discern relationships between them. This is where relations enter. A mathematical **relation** allows one to compare constants, variables, or expressions of these (or, if you prefer, concepts). Binary relations (i.e., the relation between two constants/variables/expressions or concepts) are easiest to consider, so we will restrict the discussion to the two variable case, but the idea can be generalized to an n-ary relation. Similarly, we can define orders on sets, but these admit many possibilities and are less commonly observed in political science, so we will eschew this topic as well.

A binary **relation** can be represented as an ordered pair. So, if $a \in A$ is greater than $b \in A$, we can write the relation as (a, b). When constants or variables are drawn from the integers or real numbers, though, we have more familiar notation. Integers and real numbers have natural associated orders: three is greater than two is greater than one, and so on. When one is certain of the value of a concept, as one is with a constant, then we can write $3 > 2$, $1 < 4$, and $2.5 = 2.5$. The symbols $>$, $<$, and $=$ form the familiar relations of arithmetic. When one is less sure of the values of a concept, as one is with a variable, then we also have the relations $\geq$ and $\leq$, as in $x \geq z$. Algebra, reviewed in the next chapter, deals with the manipulation of these sorts of relations.

The concept of relations is more general than these orders, however. A relation exists between two sets (or concepts) when knowing one element provides information about the other element. So, for example, in networks the relation could be "linked," while in game theory it might be "like as well as." We will explore this latter idea more in Chapter 3. While relations can be specified quite generally,[16] typically we will only be concerned with a few types of relation. Inequalities are one, and preference relations, discussed in Chapter 3, are another. The most common relation we'll use, though, is a function, which is the topic of Chapter 3. In this context we want to know the mapping between sets A and B. In other words, we want to know how the function transforms an element of A into an element of B. In this case we call A the **domain** and B the **range**. Relations (and so functions) can have various properties, some of which we discuss in Chapter 3.

[16]A relation is a mathematical object that takes as input two sets A and B (called its domain in this context) and returns a subset of $A \times B$ (called its graph in this context).

1.4.1 Why Should I Care?

Relations are important because they help us describe the mapping of values across concepts. Relations such as "greater than" and "equal to" are critical to descriptive claims about the world as well as to making theoretical claims. Further, functions—a specific type of relation—are very common in both theoretical and empirical work in political science.

1.5 LEVEL OF MEASUREMENT

We now have most of the building blocks we need to describe relationships between concepts. These in turn allow us to distinguish among different levels of measurement: **nominal**, **ordinal**, **interval**, and **ratio**. Note that though levels of measurement tend to be associated with variables, they are equally applicable and important to conceptualization.[17] We briefly discuss each level of measurement in turn.

1.5.1 Differences of Kind

In some theories all we require of our concepts is that they distinguish one type from another. That is, some concepts are about differences of kind, but not differences of degree. Concepts that identify different types but do not order them on any scale are nominal, and they require only nominal level measurement of their indicators.[18]

Nominal level measurement does not establish mathematical relationships among the values. In other words, it does not make sense to assert that a case with a nominal value of 3 is greater than one with a nominal value of 1, or that two cases with a nominal value of 2 are equal. The symbols $<$, $\leq$, $=$, $\geq$, and $>$ have no meaning for variables measured at the nominal level. Gender is a good example of a nominal level variable. When entering data for a measure of gender into a computer a researcher might assign the values of 0 and 1 (or 1 and 2) to female and male, respectively. But she might also have assigned the values -64 and $3,241$. Or she might have assigned the values 1 and 0 (or 2 and 1) to female and male, respectively. The point is that higher values do not convey any meaning: the numerical values are placeholders that indicate a difference, but the numerical values do not tell us anything meaningful.

1.5.2 Differences of Degree

At other times we are interested in differences of degree. Whether one case has more, is stronger, etc., is important to us as we define concepts and then think

[17] Students interested in an extended discussion will find Cohen and Nagel (1934, pp. 223–44) useful.

[18] The four levels of measurement—nominal, ordinal, interval, and ratio—were proposed by Stevens (1946).

about ways to measure those concepts. In such cases, nominal level concepts and measures are inadequate for they do not imply mathematical relationships among the values.

Ordinal level measurement, on the other hand, does imply mathematical relationships among the values. More specifically, the symbols $<$, $\leq$, $=$, $\geq$, and $>$ have meaning for ordinal level concepts (variables). However, the distance between any two values does not measure a constant quantity across the values the variable might take. For example, a voting scholar might be interested in people's self-placement on an ideological scale. He might put together a survey that includes a question asking people to mark themselves as far left, moderate left, middle of the road, moderate right, far right. Such a concept makes "greater than," "less than," and "equal to" distinctions. For example, we can say that moderate left is further to the left on the scale than middle of the road. And when we assign numerical values we do not have the same freedom as with a nominal measure. That is, once we have assigned two values, we are constrained on others. For example, if we assign "middle of the road" the value 3 and "far left" the value 1, then we must assign "moderate left" a value greater than 1 and less than 3. If this were a nominal level variable, then we would not be so constrained and could assign any value we wish. But ordinal variables must use numerical values that retain the order of the concept's values because the order matters in the sense that it conveys meaning. So concepts with an ordinal level of measurement have ordered values that indicate "more than" and "less than."

The next level of measurement is **interval**. This requires that the distance between values be constant over the range of values. This property is important because it makes addition and subtraction meaningful. One cannot meaningfully add or subtract variables with nominal or ordinal values because the operation does not make sense. To see that this is so, consider that we can assign any values to a binary nominal variable: $0, 1$; $1, 2$; or -64 and $3,241$. We cannot meaningfully add or subtract the values of such a variable because the values do not have meaning as numerical values. Ordinal measures, on the other hand, have meaning up to "greater than" and "less than" operations, but they also cannot be added or subtracted. If one considers the example above, we might assign the numerical values 1, 2, 3, 4, and 5 to the ideology scale, or we might assign the numerical values -3, 2, 7, 44, and 1,324. Any set of numerical values that retains the order of the concept's values is valid. The distances in the first numerical value set are constant (they are each one unit apart), but the distances in the second set vary. As such, and because both sets of values are valid, the addition and subtraction of ordinal measures do not have meaning.

Interval level measures, on the other hand, have meaningful distances between values: the intervals between numbers are constant across the range of values. Put differently, a change of $\pm x$ on the scale is the same distance regardless of where one is on the scale.

Interval levels measures may be **discrete** or **continuous**. Discrete variables with interval level measurement are integers (or natural numbers). For example, a common survey item is the feeling thermometer, which asks respondents to identify the strength of their feelings toward a politician on a scale of 0 to 100, where 0 represents extremely cold and 100 represents extremely hot (e.g., Cain, 1978; Abramowitz, 1980). Most researchers submit that the respondent recognizes that a shift of +10 points is the same anywhere on the scale.[19] That is, the distance from 0 to 5 is equivalent to the distance from 26 to 31, from 47 to 52, from 83 to 88, etc. To the extent that this is so, the measure is interval. One can meaningfully add and subtract interval level measures.

Ratio level variables are interval level variables that have a meaningful zero value. The feeling thermometer variable has a zero value, but it does not represent the absence of feeling. Instead, it represents a very strong feeling: intense dislike. So zero is not a meaningful point on the scale. As such, while we can conduct meaningful addition and subtraction operations with such variables, we cannot conduct meaningful multiplication and division operations.

The label "ratio level" comes from the fact that the same ratio at two points on the scale conveys the same meaning. This is not terribly intuitive, so let us explain. On an interval level scale any distance x between two points has the same meaning, regardless of where we are on the scale. Ratio level measurement also has this property, but it has a constant ratio property that interval level measurement lacks: the ratio of two points on the scale conveys the same meaning regardless of where one is on the scale. A good example of a ratio level scale is a public budget. Imagine that a municipal government spends four times as much on public safety as it does on public health. This is a ratio of 4:1.[20] Thus, if the city spends $4.8 million on public safety, it must spend $1.2 million on public health. Similarly, if it spends $2 million on public safety, it must spend $0.5 million on public health. Ratios can only convey meaning (i.e., measure a constant ratio) when the scale over which they are measured has a 0 value that indicates the absence (i.e., none of) whatever is being measured.

To return to the feeling thermometer example, if the value 0 represents intense negative affect (i.e., dislike), 50 indicates an absence of affect (i.e., indifference), and 100 represents intense positive affect, then 0 is not an absence of affect. Thus, it is an interval level scale, not a ratio level scale, and we cannot conclude that the first member of two pairs of respondents with scores of 20 and 10, and 50 and 25, respectively, each have twice as much affect for a candidate as the second member of each pair. However, we could rescale the feeling thermometer to make it centered on zero, perhaps assigning the value of −50 to intense negative affect, 0 to the absence of affect (or indifference), and 50 to intense positive affect. Doing so would transform the level of measurement from interval to ratio.[21]

---

[19] Note that the respondents' (implicit) beliefs about the scale of the item are important in survey research.

[20] We discuss ratios in more detail in the first section of Chapter 2. You may want to skip ahead to there if you are unfamiliar with ratios.

[21] You may be thinking that this is a trivial transformation that is not consequential, but this is not the case. To see why, try the following. Arbitrarily select a ratio—perhaps 3:1—and select two pairs of points on the transformed feeling thermometer (the one with the proper

There are lots of examples of discrete ratio level variables in political science. Political scientists are often interested in the number of events that occur, and an event count has a meaningful constant distance between values and a meaningful zero point. Thus, they are ratio variables. Examples of event counts that have been used in political science include the number of seats a party holds in parliament, the number of vetoes issued by an executive, the number of unanimous decisions by a court, and the number of wars in which a country has participated.

Thus far we have restricted our attention to discrete variables. Continuous variables have an interval or ratio level of measurement, depending on whether the value 0 represents the absence of the concept. The vast majority of (empirical) concepts that political scientists have either created or borrowed from other disciplines are discrete, but some examples of continuous measures of interest to political scientists are income and GDP.[22]

You have likely noticed that each level of measurement subsumes the levels below it. That is, ordinal level measurement is also nominal, and an interval measure has ordinal and nominal properties. This suggests that whenever we have a concept at a high level of measurement we can reconceptualize and remeasure it at a lower level of measurement should we have cause to do so.

Some people mistakenly view the hierarchy of the levels of measurement as a means to judge the heuristic value of concepts. This is an error. Concepts can be evaluated on their clarity (vague concepts have little heuristic value), and one can make normative judgments about concepts (e.g., freedom, peace, order), but all sufficiently clear concepts are merely inputs to specific theories, and theories, not their concepts, should be judged. A proper discussion of this issue is beyond the scope of this book, but it is important to recognize that a nominal conceptualization may yield insights that a ratio conceptualization would miss and vice versa. Put differently, it would be an error to judge the levels of measurement as an ordinal scale with respect to their value to causal theory: it is nominal.

1.5.3 Why Should I Care?

Recognizing whether one is thinking about differences of kind (nominal) or degree (ordinal, interval, or ratio) is critical. If one is thinking about differences of degree, then how precise are those differences? Without a firm grasp on levels of measurement one cannot be precise about one's concepts, much less one's measures of one's concepts.

---

ratio scale where -50 is intense dislike, 0 is indifference, and 50 is strong positive affect) that have that ratio. Now transform the scale to the actual feeling thermometer (the one with the range from 0 to 100). Recalculate the ratios. They are different, right? The two scales do not produce the same ratio levels, and that means that one of them preserves ratios and the other does not. The one with the meaningful zero is the only scale that produces meaningful ratios. For a more detailed explanation, see Stevens (1946).

[22] If one rounds either to dollars, thousands of dollars, etc., then the values are integers (or natural numbers) and the measure is discrete.

1.6 NOTATION

Here we list, and in some cases briefly describe, common notation. This section is one you will likely refer to from time to time, but not everything might be clear now. Also, as a reference section it is heavier on the math and lighter on the intuition. It is important to read it once now, but if you find yourself unclear on some notation later, please refer back to this section. To make reference easier, we begin with the summary Table 1.2.

Operators take many forms, and are commonly used. We have already discussed some: $+$, $-$, $\times$, $/$, x^n, $\sqrt[n]{x}$, $\sum$, $\prod$, $!$. Some of these have multiple ways to represent them, others mean multiple things depending on context. For example, there are several ways to represent multiplication: $a \times b \times c = a * b * c = a \cdot b \cdot c = abc$. Of course, as we have seen, $\times$ can also mean a Cartesian product when applied to sets. Both $/$ and $\div$ mean divide; the mod operator, written $8 \bmod 3$, means divide the first number by the second, and report the remainder: $8 \bmod 3 = 2$.

One can also use the product operator, $\prod$, to represent the product of a, b, and c: $\prod_a^c$.

> One reads that as *the product of a through c*.

More typically, the product operator is used by indexing a variable (this is accomplished by adding a subscript: x_i) and writing: $\prod_{i=k}^{l} x_i$.

> One reads that as *the product of x_i over the range from $i = k$ through $i = l$*.

When the product operator is used in an equation that is set apart from the text, it looks like this:

$$\prod_{i=k}^{l} x_i = x_k \times \ldots \times x_l.$$

The "..." here signals the reader to assume all interim values are included in the product. When used at the end of a list, e.g., $1, 2, 3, \ldots$, "..." signifies that the list (or product or sum) goes on indefinitely. In these cases you may also see ∞ as an end to the sequence instead, e.g., $1, 2, 3, \ldots, \infty$; ∞ is the symbol for infinity. In other words, ... means continue the progression until told to stop.

The summation operator, $\sum$, can be used to represent the addition of several numbers. For example, if we want to add together all members of a set indexed by i, then we can write: $\sum_i$. One reads that as *the sum over i*. You will also see summation represented over a range of values, say from value k through value l: $\sum_{i=k}^{l} x_i$.

> One reads that as *the sum of x_i over the range from $i = k$ through $i = l$*.

Table 1.2: Summary of Symbols and Notation

Symbol	Meaning
$+$	Addition
$-$	Subtraction
$*$ or $\times$ or $\cdot$	Multiplication
$/$ or $\div$	Division
$\pm$	Plus or minus
x^n	Exponentiation ("to the nth power")
$\sqrt[n]{x}$	Radical or nth root
$!$	Factorial
∞	Infinity
$\sum_{i=k}^{l} x_i$	Sum of x_i from index $i = k$ to $i = l$
$\prod_{i=k}^{l} x_i$	Product of x_i from index $i = k$ to $i = l$
$\ldots$	Continued progression
$\frac{d}{dx}$	Total derivative with respect to x
$\frac{\partial}{\partial x}$	Partial derivative with respect to x
$\int dx$	Integral over x
$\cup$	Set union
$\cap$	Set intersection
$\times$	Cartesian product of sets
$\setminus$	Set difference
A^c	Complement of set A
$\emptyset$	Empty (or null) set
$\in$	Set membership
$\notin$	Not member of set
$\mid$ or $:$ or $\ni$	Such that
$\subset$	Proper subset
$\subseteq$	Subset
$<$	Less than
$\leq$	Less than or equal to
$=$	Equal to
$>$	Greater than
$\geq$	Greater than or equal to
$\neq$	Not equal to
$\equiv$	Equivalent to or Defined as
$f()$ or $f(\cdot)$	Function
$\{\ \}$	Delimiter for discrete set
$(\ )$	Delimiter for open set
$[\ ]$	Delimiter for closed set
$\forall$	For all (or for every or for each)
$\exists$	There exists
$\Rightarrow$	Implies
$\Leftrightarrow$	If and only if
$\neg C$ or $\sim C$	Negation (not C)

Set apart from the text in an equation, the summation operator looks like this:

$$\sum_{i=k}^{l} x_i = x_k + \ldots + x_l.$$

The exponential operator, x^n (read "x to the nth power," or "x-squared" when $n = 2$ and "x-cubed" when $n = 3$), represents the power to which we raise the variable, x. The root operator, $\sqrt[n]{x}$ (read "the nth root of x," or "the square root of x" when $n = 2$ or "the cube root of x" when $n = 3$), represents the root of x.

Factorial notation is used to indicate the product of a specific sequence of numbers. Thus, $n! = n \times (n-1) \times (n-2) \ldots \times 2 \times 1$. So $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$, and $10! = 10 \times 9 \times \ldots 3 \times 2 \times 1 = 3,628,800$. This notation is especially useful for calculating probabilities.

You may not be familiar with some of the operators used in calculus. The derivative of x with respect to t is represented by the operator $\frac{dx}{dt}$. The operator ∂ indicates the partial derivative, and $\int$ indicates the integral. These will be the focus of Parts II and V of this book.

Though it's not an operator, one more symbol is useful to mention here: $\pm$. Read as "plus or minus," this symbol implies that one cannot be sure of the sign of what comes next. For example, $\sqrt{4} = \pm 2$, because squaring either 2 or -2 would produce 4.

Sets, as we have seen, have a good deal of associated notation. There are the set operators $\cap$, $\cup$, $\times$, and $\setminus$, plus the complement of A (A^c or A'). There are also the empty set $\emptyset$, set membership $\in$, set nonmembership $\notin$, proper subset $\subset$, and subset $\subseteq$. To these we add $|$, : , or $\ni$, which are each read as "such that." These are typically used in the definition of a set. For example, we define the set $A = \{x \in B | x \leq 3\}$, read as "the set of all x in B such that x is less than or equal to 3." In other words, the $|$ indicates the condition that defines the set. It serves the same purpose in conditional probabilities ($P(A|B)$), as we will see in Part III of the book. Sets also make use of delimiters, described below.

Relations include $<$, $\leq$, $=$, $\geq$, $>$. They also include $\neq$, which means "not equal to," and $\equiv$, which means "exactly equivalent to" or, often, "defined as." Relations between variables or constants typically have a left-hand side, to the left of the relation symbol, and a right-hand side, to the right of the relation symbol. These are often abbreviated as LHS and RHS, respectively. Functions are typically written as $f()$ or $f(\cdot)$, both of which imply that f is a function of one or more variables and constants. The "$\cdot$" here is a placeholder for a variable or constant; do not confuse it with its occasional use as a multiplication symbol, which occurs only when there are things to multiply.

Delimiters are used to indicate groups. Sometimes the groups are used to identify the order of the operations that are to be performed: $(x + x^2)(x - z)$. One performs the operations inside the innermost parentheses first and then moves outward. Square braces and parentheses are also used to identify closed and open sets, respectively. The open set (x_1, x_n) excludes the endpoint values

Table 1.3: Greek Letters

Upper-case	Lower-case	English	Upper-case	Lower-case	English
A	α	alpha	N	ν	nu
B	β	beta	Ξ	ξ	xi
Γ	γ	gamma	O	o	omicron
Δ	δ	delta	Π	π	pi
E	ϵ	epsilon	P	ρ	rho
Z	ζ	zeta	Σ	σ	sigma
H	η	eta	T	τ	tau
Θ	θ	theta	Υ	υ	upsilon
I	ι	iota	Φ	ϕ	phi
K	κ	kappa	X	χ	chi
Λ	λ	lambda	Ψ	ψ	psi
M	μ	mu	Ω	ω	omega

x_1 and x_n, whereas the closed set $[x_1, x_n]$ includes the endpoint values x_1 and x_n. Curly braces are used to denote set definitions, as above, or discrete sets: $\{x_1, x_2, \ldots, x_n\}$. Parentheses are also often used for ordered pairs or n-tuples, as we have seen; for example, $(2, 3, 1)$. They are also often used in vectors, which have a similar meaning. Both parentheses and square braces are used interchangeably to indicate the boundaries of matrices. We will discuss both vectors and matrices in Part IV of the book.

Proofs, the topic of the next section, have their own notation, which may pop up in other sections as well. The symbol $\forall$ means "for all," so $\forall x \in A$ means the associated statement applies for all x in the set A. The symbol $\exists$ means "there exists," typically used in the context of $\exists$ some $x \in A$ such that $x < 3$. The symbol $\Rightarrow$ is read as "implies" and is used as $C \Rightarrow D$, which means that whenever statement C is true, D is too. One can also use the reverse, $C \Leftarrow D$, which means that C is true if D is true. The symbol $\Leftrightarrow$ means that both implications are true and is read as "if and only if," so $C \Leftrightarrow D$ means that C is true if D is true, and only if D is true. In other words, C and D are equivalent statements. The symbol $\neg$ denotes negation, so $\neg C$ means statement C is not true. You will also sometimes see $\sim C$ used to mean C is not true.

People sometimes use **Greek letters** to represent variables, particularly in formal theory; they are often used to represent constants (aka parameters) in statistical analysis. Table 1.3 lists the Greek alphabet. If you have never encountered the Greek alphabet you may want to make a copy of this page, cut out the table, and tape it to the wall where you study for this and other courses that use math. Or just save it to your preferred portable electronic device.

1.6.1 Why Should I Care?

Notation that you cannot read is a serious stumbling block to understanding!

1.7 PROOFS, OR HOW DO WE KNOW THIS?

As we progress through this book, we will offer up a great many pieces of information as fact, often without explaining how we knew they were true. As noted in the preface to this book, we do this in order to focus on intuition rather than mathematical formalism. However, it is certainly fair to wonder—more than fair, really—how one comes to these conclusions. The answer, as we discuss briefly here, is that they have been proven to be true.

How does this work? Mathematics is not an empirical science; there are no experiments, and no data except insofar as experience shapes the thought of mathematicians. Rather, the progress of math begins with axioms and **assumptions**, which are stated up front with clarity and taken to be true.[23] One then conjectures a **proposition**, which is just a statement that is thought to be true given the assumptions made. From these assumptions, along with any previously proved theorems, one deductively proves, or disproves, the proposition. A proven proposition is often referred to as a **theorem**, unless it is of little interest in and of itself and is intended to be used only as a stepping stone, in which case it is called a **lemma**. A **corollary** is a type of proposition that follows directly from the proof of another proposition and does not require further proof. You will see assumptions and propositions commonly in pure and applied game theory, and lemmas, theorems, and corollaries somewhat less commonly. Propositions, though deductively derived, are often empirically testable given appropriate measures for the variables used in the proposition. In other words, a proposition might state that y is increasing in x_1 and decreasing in x_2. To test this empirically, one needs measures for y, x_1, and x_2. In some scientific fields it is common to distinguish propositions from hypotheses, with the former referring to statements of expected relationships among concepts and the latter referring to expected relationships among variables. In such contexts propositions are more general statements than hypotheses. At present, these distinctions are not widely used among political scientists.

It is not difficult to make assumptions, though learning to specify them clearly and to identify the implicit assumptions you may be making takes practice. Nor is it difficult to state propositions that may be true, though similar caveats apply. The tricky part is in proving the proposition. There is no one way to prove all propositions, though the nature of the proposition can suggest the appropriate alternative. We will consider a few commonly observed methods here, but this is far from a complete accounting.

We begin by considering four statements: A, B, C, D. A statement can be anything, e.g., A could be $x < 3$ or "all red marbles are in the left urn" or "democracies are characterized primarily by elections." Let's assume that A and B are assumptions. We take them to be true at the start of our proof and

[23] Political scientists rarely specify axioms, which tend to be more significant and wide-ranging assumptions than what are called simply assumptions. The following discussion uses terms as they are commonly observed in political science, which may elide mathematical nuance.

will not deduce them in any way from other statements. Of course, if they are not empirically true, then our conclusions may very well be incorrect empirically, but, as you can guess by the repeated use of the word "empirically," this is an empirical question and not a mathematical one. Let's further assume that C is an interim statement—that is, a deduced statement that is not our intended conclusion—and that D is that conclusion. Thus our goal is to derive D from A and B. This is the general goal of mathematical proofs.

More precisely, in this case we are seeking to show that A and $B \Rightarrow D$ (A and B imply D). This is a sufficiency statement: A and B are *sufficient* to produce D. We also can call this an *if* statement: D is true if A and B are true. This is not the only possible implication we could have written (implications are just a type of mathematical statement). We could instead have stated that A and $B \Leftarrow D$ (A and B are implied by D). This is a statement of necessity: A and B are *necessary* to produce D, since every time D is true, so are A and B. We can also call this an *only if* statement: D is true only if A and B are. Take a moment to think about the difference between these two ideas, as it is fairly central to understanding theory in political science, and it is not always obvious how different the statements are.

Ready? There is also a third common implication we could have written, a necessary *and* sufficient statement: A and $B \Leftrightarrow D$. This is also called an *if and only if* statement, as D is true if and only if A and B are true. In other words, A and B are entirely equivalent logically to D, and one can replace one statement with the other at will. This is one way one uses existing theorems to help in new proofs, by replacing statements with other statements proven to be equivalent. (One can also use if or only if propositions on their own in new proofs.)

In addition to using existing theorems, pretty much any mathematical procedure accepted as true can be used in a proof. We'll cover many in this book, but the most basic of these may be the tools of formal **logic**, which has much in common with set theory. Negation of a statement is much the same as the complement of a set. For example, you cannot be both true and not true, nor can you be both in and outside a set. You can also take the equivalent of a union and an intersection of sets for statements; these are called disjunction and conjunction, or, in symbols, *or* ($\vee$) and *and* ($\wedge$), respectively. Note that the *and* symbol looks like the intersection symbol. This is not accidental—*and* means that both statements are true, which is like being in both sets, which is like the intersection of the sets. Likewise, *or* means that at least one statement is true, which is like being in either set, which is like the union between the sets. Let's call a compound statement anything that takes any two simpler statements, such as A and B, and combines them with a logical operator, such as $\neg, \vee$, or $\wedge$. We can therefore write the implication we're trying to prove as $A \wedge B \Rightarrow D$.

De Morgan's laws prove handy for manipulating both sets and logical statements.[24] We'll present these in terms of logical statements, but they are true for sets as well after altering the notation. The best way to remember them is that the negation of a compound statement using *and* or *or* is the compound statement in which the *and* is switched for *or*, or vice versa, and each of the simpler statements is negated. So, for example, $\neg(A \wedge B)$ is $(\neg A) \vee (\neg B)$ and $\neg(A \vee B)$ is $(\neg A) \wedge (\neg B)$. In words, if both statements aren't true, then at least one of them must be false. Similarly, if it's not the case that at least one statement is true, then both statements are false.

We can use logic to obtain several important variants of our implications that might be useful. A negated implication just negates all the statements that are part of the implication. So the negation of our implication becomes $\neg(A \wedge B) \Rightarrow \neg D$, which by De Morgan's law is $(\neg A) \vee (\neg B) \Rightarrow \neg D$. Even when the statement is true, the negation might not be. Having two democracies may mean you're at peace (for the sake of this argument), but letting at least one of them not be a democracy does not automatically imply war.

The **converse** of an implication switches a necessary statement to a sufficient one, or vice versa. Thus the converse of A and $B \Rightarrow D$ is A and $B \Leftarrow D$ or $D \Rightarrow A$ and B. As noted above, just because an implication is true does not mean the converse is true—something may be necessary without being sufficient. However, negating the converse, called taking the contrapositive, *does* always yield a true statement. The **contrapositive** of our implication is $(\neg A)\vee(\neg B) \Leftarrow \neg D$, or, as it's more typically written, $\neg D \Rightarrow (\neg A) \vee (\neg B)$. If a pair (dyad) of democracies never experiences war, then having a war (the opposite of peace) means that at least one of the pair is not a democracy.

Okay, back to our proof. Proofs are sometimes classed into broad groups of direct and indirect proofs. **Direct proofs** use deduction to string together series of true statements, starting with the assumptions and ending with the conclusion. In addition to the construction of a string of arguments, direct proofs commonly observed in formal theory include proof by exhaustion, construction, and induction. Let us see briefly how these work, starting with a **general deductive proof**.

Let A be the statement that $x \in \mathbb{Z}$ is even, and B be the statement that $y \in \mathbb{Z}$ is even, and D, which we're trying to prove, be the statement that the product xy is even. Well, if x and y are even (our assumptions), then they can be written as $x = 2r$ and $y = 2s$ for some $r, s \in \mathbb{Z}$. (Here we've used the definition of even.) In this case, we can write $xy = (2r)(2s) = 4rs$, which is our new statement C. Since $4rs = 2(2rs)$, xy is even (again using the definition of even), thus proving D. Now we know that the product of any two even integers is also even, and we could use this knowledge in further, more complex proofs.

Proof by exhaustion is similar, save that you also break up the problem into exhaustive cases and prove that your statement is true for each case. This comes up often in game theory as there will be different regions of the parameter space that may behave differently and admit different solutions. (The parameter space

[24] See `http://en.wikipedia.org/wiki/De_Morgan_laws`.

is the space, in the sense of a set with a measure, spanned by the parameters. We will discuss this concept more in Part III of the book.)

Proof by construction is similarly straightforward, and can be useful when trying to show something like existence: if you can construct an example of something, then it exists.

Proof by induction is a bit different and merits its own example. It is generally useful when you would like to prove something about a sequence (we cover sequences in Chapter 4) or a sequence of statements. It consists of three parts. First, one proves the base case, which in this example is the first element in the sequence. Second, one assumes that the statement is true for some n (the inductive hypothesis). Third, one proves that the statement is true for $n+1$ as well (the inductive step). Thus, since the base case is true and one can always go one further in the sequence and have the statements remain true, the entire sequence of statements is true.[25] Let's see how this works with an example: show that $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. We basically need to show this is true for each n, but since they occur in sequence, we'll use induction rather than exhaustion (which wouldn't be appropriate, given that the sequence is infinite anyway). First we try the base case, which is for $n = 1$. We can check this: $\sum_{i=1}^{1} i = 1 = \frac{1(2)}{2} = \frac{1(1+1)}{2}$. So the base case is true. Now we assume, somewhat counterintuitively, the statement that we're trying to prove: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$. Finally, we show it remains true for $n+1$, so we need to prove that $\sum_{i=1}^{n+1} i = \frac{(n+1((n+1)+1)}{2}$, where we've replaced n in the right-hand side of the statement we're trying to prove with $n+1$. The sum in the left-hand side of this is $\sum_{i=1}^{n} i + (n+1)$, where we've just split the sum into two pieces. The first piece equals $\frac{n(n+1)}{2}$ by step two in our proof. So now we have $\frac{n(n+1)}{2} + (n+1) = \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{n(n+1)+2(n+1)}{2} = \frac{(n+1)(n+2)}{2}$. This is just what we needed to show, so the $n+1$ inductive step is true, and we've proved the statement.

Indirect proofs, in contrast, tend to show that something must be true because all other possibilities are not. **Proof by counterexample** and **proof by contradiction** both fall into this category. Counterexamples are straightforward. If the statement is that $A \wedge B \Rightarrow D$ and A and B are both true, then a single counterexample of $\neg D$ is sufficient to disprove the proposition. Proof by contradiction has a similar intent, but instead of finding a counterexample one starts by assuming the statement one is trying to prove is actually false, and then showing that this implies a contradiction. This proves the proposition because if it cannot be false, then it must be true. Although it may seem counterintuitive, proof by contradiction is perhaps the most common type of proof, and is usually worth trying first. Proving the contrapositive, since it indirectly

[25]Though this method of proof is called mathematical induction, it's important to note that it is a deductive method of theory building, not an inductive one. That is, it involves making assumptions and deducing conclusions from these, not stating conclusions derived from a series of statements that may only be probabilistically linked to the conclusion, as in inductive reasoning.

also proves the statement, as they are equivalent, is sometimes also considered an indirect proof, though it seems pretty direct to us.

1.8 EXERCISES

1.8.1 Constants and Variables and Levels of Measurement

1. Identify whether each of the following is a constant or a variable:
   a) Party identification of delegates at a political convention.
   b) War participation of the Great Powers.
   c) Voting record of members of Congress relative to the stated position of the president.
   d) Revolutions in France, Russia, China, Iran, and Nicaragua.
   e) An individual voter's vote choice in the 1992 presidential election.
   f) An individual voter's vote choice in the 1960–1992 presidential elections.
   g) Vote choice in the 1992 presidential election.
2. Identify whether each of the following is a variable or a value of a variable:
   a) The Tonkin Gulf Crisis.
   b) Party identification.
   c) Middle income.
   d) Exports as a percentage of GDP.
   e) Republican.
   f) Female.
   g) Veto.
   h) Ethnic fractionalization.
   i) International crisis.
3. Identify whether each of the following indicators is measured at a nominal, ordinal, interval, or ratio level. Note also whether each is a discrete or a continuous measure:
   a) Highest level of education as (1) some high school, (2) high school graduate, (3) some college, (4) college graduate, (5) postgraduate.
   b) Annual income.
   c) State welfare expenditures, measured in millions of dollars.
   d) Vote choice among Bush, Clinton, and Perot.
   e) Absence or presence of a militarized interstate dispute.
   f) Military personnel, measured in 1,000s of persons.
   g) The number of wars in which countries have participated.

1.8.2 Sets, Operators, and Proofs

4. As a brief illustration of one use of set theory, consider the following question: given three parties in a legislature with a supermajority rule required to pass a bill, what is the likely outcome of a given session? We can use set theory and some rational choice assumptions to get a pretty good handle on that question. Assume that no party has enough seats to pass the bill by itself and that all three parties prefer some outcome other than the status quo. For concreteness, let's define two dimensions over which to define policy: guns (i.e., defense spending) and butter (i.e., health, education, and welfare spending). We can now create a two-dimensional space where spending on guns is plotted on the vertical axis and spending on butter is plotted on the horizontal axis. Take out a sheet of paper and draw this. Let the axes range from 0% of the budget, marked where the axes intersect, to 100% of the budget, marked as the maximum value on each axis. Connect the two maximum values with a straight line. You now have a triangle, and the legislature cannot go outside the triangle: the line you just drew represents spending the entire budget on some mix of guns and butter. Let's assume that the legislators want to spend some money on non-guns and non-butter, and thus both parties' most preferred combination of guns and butters is somewhere inside the budget constraint. Pick some point inside the budget constraint and mark it as the status quo. Now select a most preferred combination for each party and mark each as Party 1, Party 2, and Party 3. Finally, pick a fifth point and label it a bill. Make a conjecture on whether the bill will pass or whether the status quo will be sustained. (For now this is just a conjecture, but we'll return to this in the exercises to Chapter 3, so save your answer.)

5. Let $A = \{1, 5, 10\}$ and $B = \{1, 2, \ldots, 10\}$.

   a) Is $A \subset B$, $B \subset A$, both, or neither?
   b) What is $A \cup B$?
   c) What is $A \cap B$?
   d) Partition B into two sets, A and everything else. Call everything else C. What is C?
   e) What is $A \cup C$?
   f) What is $A \cap C$?

6. Write an element of the Cartesian product $[0, 1] \times (1, 2)$.

7. Prove that $\sqrt{2}$ is an irrational number. That is, show that it cannot be written as the ratio of two integers, p and q.

8. Prove that the sum of any two even numbers is even, the sum of any two odd numbers is even, and the sum of any odd number with any even number is odd.

Chapter Two

Algebra Review

Of all the chapters in this book, this is the one most safely skipped. Most of this chapter is taken up by a review of arithmetic and algebra, which should be familiar to most readers. If you feel comfortable with this material, skip it. If it is only vaguely familiar, don't. The third section briefly discusses the utility of computational aids for performing calculations and checking work.

2.1 BASIC PROPERTIES OF ARITHMETIC

There are several properties of arithmetic that one uses when simplifying equations. These arise from the real numbers or integers for which the variables stand. In other words, because the variables we use in political science generally take values in $\mathbb{R}$ or $\mathbb{Z}$, these five properties generally apply. This will be true nearly throughout the book; however, in Part IV we will see that matrix variables can fail to commute under multiplication, for example, and do not always possess multiplicative inverses. But for variables that stand for real numbers or integers, these properties will always hold. Most of these are expressed in terms of addition and multiplication, but the first three properties apply to subtraction and division, respectively, as well, except for division by zero.

The **associative properties** state that $(a+b)+c = a+(b+c)$ and $(a\times b)\times c = a \times (b \times c)$. In words, the properties indicate that the grouping of terms does not affect the outcome of the operation.

The **commutative properties** state that $a + b = b + a$ and $a \times b = b \times a$. In words, the properties claim that the order of addition and multiplication is irrelevant.

The **distributive property** states that $a(b + c) = ab + ac$. In words, the property says that multiplication distributes over addition (and subtraction).

The **identity properties** state that there exists a zero such that $x + 0 = x$ and that there exists a one such that $x \times 1 = x$. In other words, there exist values that leave x unchanged under addition and multiplication (and subtraction and division, respectively, as well).

The **inverse property** states that there exists a $-x$ such that $(-x) + x = 0$. In other words, there exist values that when added to any x produce the identity under addition. We might also consider an inverse under multiplication, x^{-1}, such that $(x^{-1}) \times x = 1$. The existence of this inverse is a property of the real numbers (and the rational numbers), but not the integers, so one must be careful. For example, if $x = 2$, then $x^{-1} = 0.5$ in the real numbers, but no integer multiplied by two equals one. Whether or not an inverse exists will depend, therefore, on the set of values the variable can take.

It is useful to recall at this stage that division by zero is undefined. The expression $x/0 = \infty$ is true for any $x \neq 0$, but is completely undefined for $x = 0$. Other useful facts include that $x = 1x = x^1 = 1x^1$, and that $x^0 = 1$. Recall also that multiplication by a variable with a negative value changes the sign of the product: $-1 \times x = -x$. The product of two terms with negative signs is positive: $(-x) \cdot (-y) = xy$.

2.1.1 Order of Operations

The order of operations is also important and can trip people up. In arithmetic and algebra the order of operations is parentheses, exponents, multiplication, division, addition, subtraction. A common mnemonic device people use to memorize order of operations is PEMDAS, or <u>P</u>lease <u>E</u>xcuse <u>M</u>y <u>D</u>ear <u>A</u>unt <u>S</u>ally.

2.1.2 Ratios, Proportions, and Percentages

Ratios, proportions, and percentages sometimes give people trouble, so let's briefly review those. The **ratio** of two quantities is one divided by the other $\frac{x}{y}$ is the ratio of x to y. Ratios are also written as $x : y$. Keep in mind that one can only take the ratio of two variables measured at a ratio level of measurement (i.e., there is a constant scale between values, and a meaningful zero). Though a ratio may be negative, we typically consider ratio variables that range from 0 to ∞. To get this, we take the absolute value of the ratio, denoted $\left|\frac{x}{y}\right|$. All this does is turn any negative number positive. As an example, international relations scholars are often interested in the ratio of military power between two countries (e.g., Organski and Kugler, 1981).

The **proportion** of two variables, on the other hand, is the amount one variable represents of the sum of itself and a second variable: $\left|\frac{x}{x+y}\right|$. A proportion ranges from a minimum of 0 to a maximum of 1. Students of budgetary politics are often interested in the proportion of expenditures that is spent in a given category (e.g., health and welfare, pork barrel politics, defense spending; see Ames, 1990).

The **percentage** one variable represents of a total is the proportion represented over the range from 0 to 100. In other words, the percentage is a linear transformation of the proportion $\left|\frac{x}{x+y}\right| \times 100\%$. Many people find a percentage representation more intuitive than a proportion representation, but they provide the same information.

You will also encounter the **percentage change** in a variable, which is calculated as $\frac{(x_{t+1} - x_t)}{x_t}$, where the subscript t indicates the first observation and the subscript $t + 1$ indicates the second observation. For instance, according to the Center for Defense Information's *Almanac*, the United States spent \$75.4 billion for military personnel wages in 2001 and an estimated \$80.3 billion in

2002. The expenditures in 2002 represented a 6.5% increase over 2001 expenditures: $\frac{(80.3-75.4)}{75.4} \simeq 6.5\%$. Note that the percentage change can range from $-\infty$ to ∞.

2.1.3 Why Should I Care?

You care about these properties because you need to know them to follow along. People who use mathematics to communicate their ideas, whether in formal theory or statistics, assume that you can do the operations allowed by these properties. They often "skip steps" when writing down manipulations and expect you to do them in your head. If you cannot do them, you will get lost.

2.2 ALGEBRA REVIEW

This section reviews the most common algebraic manipulations you will encounter. Most of you will be familiar with these; the trick is trying to minimize errors, which are easy to make. We note some common errors to avoid.

2.2.1 Fractions

Many students find fractions the most frustrating part of algebra. People generally find whole numbers more intuitive than fractions, and that makes calculations with fractions more difficult to perform. As such, whenever possible it is best to convert fractions to whole numbers. Recall that the number on the top of a fraction is the **numerator** and the number on the bottom of a fraction is the **denominator**

$$\frac{Numerator}{Denominator}.$$

Thus, one can convert to a whole number whenever the denominator divides evenly into the numerator.

Many people find mixed numbers such as $2\frac{3}{4}$ even more frustrating. To convert these mixed numbers to fractions, follow these two steps. First, multiply the denominator of the fraction by the whole number (i.e., multiply 4×2, which equals 8). Second, take this product and add it to the numerator and place that sum over the original denominator (add 8 to 3, which equals 11, and place that over 4 for the final fraction $\frac{11}{4}$). These two quantities are equivalent.

Two common algebraic manipulations relating to fractions that often trouble students are cancellations and adding fractions.

2.2.1.1 Cancellations

The reason we want to reduce fractions is to make them easier to use (if the fraction can be converted to a whole number, this is ideal). For example, $\frac{10x}{2}$ can be reduced to $5x$. One that you might encounter in game theory could look like this: $\frac{7+3x}{2x}$.

One of the most common mistakes made is to cancel the xs and simplify $\frac{7+3x}{2x}$ to $\frac{10}{2}$, and then simplify this quantity to 5. However, $7+3x \neq 10x$, so $\frac{7+3x}{2x} \neq 5$.

In this example $\frac{7+3x}{2x}$ can be simplified to $\frac{7}{2x} + \frac{3x}{2x}$. The fraction $\frac{7}{2x}$ is in its simplest form. The fraction $\frac{3x}{2x}$ can be simplified to $\frac{3}{2}$, as long as $x \neq 0$.[1]

Therefore, $\frac{7+3x}{2x} = \frac{7}{2x} + \frac{3}{2}$.

2.2.1.2 Adding Fractions

Adding or subtracting fractions can be a bit frustrating as they do not follow the same rules as whole numbers. More specifically, you can only add the numerators of two or more fractions when the denominators of each fraction are the same (i.e., you cannot add fractions with different denominators). You can add $\frac{4}{\beta} + \frac{\alpha}{\beta}$, which equals $\frac{4+\alpha}{\beta}$. When two fractions have different denominators, such as $\frac{\beta}{4} + \frac{\alpha}{2}$, one must transform one or both of the denominators to make addition possible: the numerators of all fractions can be added once their denominators are made equal.

To pursue the above example, $\frac{\beta}{4} + \frac{\alpha}{2}$, if we multiply $\frac{\alpha}{2}$ by $\frac{2}{2}$ (which equals one; you can always multiply by things equal to one, or add things equal to zero because of the identity property), it becomes $\frac{2\alpha}{4}$. Since the two fractions now have the same denominator, we can add their numerators:[2] $\frac{\beta}{4} + \frac{2\alpha}{4} = \frac{2\alpha+\beta}{4}$.

Unlike addition, multiplication does *not* require a common base, and one does multiply both numerator and denominator: $\frac{2}{3} \times \frac{1}{4} = \frac{2}{12} = \frac{1}{6}$.

Another common mistake people make when adding fractions is to assume that all aspects of fractions follow the same rules of addition. For example, they assert that $\frac{1}{\Delta+\Theta}$ is equal to $\frac{1}{\Delta} + \frac{1}{\Theta}$. It is not. To see why this is so, let's add real numbers to the expression. If we substitute 2 for Δ and 1 for Θ and sum the denominator, we get $\frac{1}{2+1}$, which is equal to $\frac{1}{3}$. If we split the fraction improperly, including the numerator over both parts of the denominator as above, we will conclude that $\frac{1}{2+1} = \frac{1}{2} + \frac{1}{1}$, which equals $1\frac{1}{2}$, or 1.5,[3] not $\frac{1}{3}$.

2.2.2 Factoring

Factoring involves rearranging the terms in an equation to make further manipulation possible or to reveal something of interest. The goal is to make the expression simpler. One uses the properties described above rather extensively when factoring.

A standard algebraic manipulation involves combining like terms in an expression. For example, to simplify $\delta + \delta^2 + 4\delta - 6\delta^2 + 18\delta^3$, we combine all like terms. In this case we combine all the δ terms that have the same exponent, which gives us $18\delta^3 - 5\delta^2 + 5\delta$.

[1] Remember, anything divided by itself is one, and anything multiplied by one equals itself.

[2] Note that we do **not** take the sum of the denominators. One only adds the numerators.

[3] If you have forgotten how to convert fractions into decimals, the solution is to do the division implied by the fraction (you can use a calculator if you wish): $\frac{1}{2} = 0.5$.

Another standard factoring manipulation involves separating a common term from unlike ones. We first establish what we want to pull out of the equation, then apply the distributive property of multiplication in reverse. For example, we might want to pull x out of the following: $3x + 4x^2 = x(3 + 4x)$. Another example is $6x^2 - 12x + 2x^3 = 2x(3x - 6 + x^2)$.

A more complex example is $12y^3 - 12 + y^4 - y$.

We can factor 12 out of the first two terms in the expression and y out of the next two terms.

The expression is then $12(y^3 - 1) + y(y^3 - 1)$, which can be regrouped as $(12 + y)(y^3 - 1)$.

2.2.2.1 Factoring Quadratic Polynomials

Quadratic polynomials are composed of a constant and a variable that is both squared and raised to the power of one: $x^2 - 2x + 3$, or $7 - 12x + 6x^2$.[4] Quadratic polynomials can be factored into the product of two terms: $(x \pm ?) \times (x \pm ?)$, where you need to determine whether the sign is $+$ or $-$, and then replace the question marks with the proper values.

Hopefully, it is apparent that one can multiply many products of two sums or two differences to get a quadratic polynomial; this is the reverse of factoring.[5]

2.2.2.2 Factoring and Fractions

We can also reduce fractions by factoring. Consider the fraction $\frac{x^2-1}{x-1}$. We can factor the numerator $x^2 - 1 = (x + 1)(x - 1)$. We can thus rewrite the fraction as follows

$$\frac{x^2 - 1}{x - 1} = \frac{(x + 1)(x - 1)}{x - 1}.$$

The term $x - 1$ is in both the numerator and the denominator and thus (as long as $x \neq 1$) cancels out, leaving $x + 1$. Thus, $\frac{x^2-1}{x-1} = x + 1$ for $x \neq 1$.

This factoring need not be accomplished in one step. Consider the expression

$$\frac{3\lambda^4 + 3\lambda^3 - 6\lambda^2}{6\lambda^2 + 12\lambda}.$$

First, we can factor out the common factor from both the numerator and denominator. All of the terms in the numerator are multiples of $3\lambda^2$ and both of the terms in the denominator are multiples of 6λ. This yields

[4] Another way of putting this is that a quadratic polynomial is a second-order polynomial in a single variable x. We discuss polynomial functions in the next chapter. Finally, given that the Latin prefix *quadri* is associated with four, you may be wondering why *quadratic* is used to describe equations with a term raised to the power of two. The reason is that the Latin term *quadratum* means "square." So an equation with a variable that is squared is a quadratic equation (Weisstein, N.d.).

[5] Note that this is true of some, *but not all*, products of two sums or two differences.

$$\frac{3\lambda^2(\lambda^2+\lambda-2)}{6\lambda(\lambda+2)}.$$

Next, we factor the quadratic polynomial in the numerator to get

$$\frac{3\lambda^2(\lambda+2)(\lambda-1)}{6\lambda(\lambda+2)}.$$

Then we factor out like terms. Both the numerator and denominator have $\lambda+2$, so (as long as $\lambda \neq -2$) they cancel out, leaving

$$\frac{3\lambda^2(\lambda-1)}{6\lambda}.$$

Finally, 3λ can be canceled (as long as $\lambda \neq 0$) from both the numerator and the denominator, leaving the expression in its simplest form

$$\frac{\lambda(\lambda-1)}{2}.$$

2.2.3 Expansion: The FOIL Method

Sometimes we need to simplify a complex expression. At other times we need to expand a simple expression. Here is a pop quiz:

Does $(\delta+\gamma)^2 = \delta^2+\gamma^2$?

The answer: no.

Why? The expression $(\delta+\gamma)^2 = (\delta+\gamma)(\delta+\gamma)$. This can then be expanded using the FOIL method. The expanded expression is $\delta^2+2\delta\gamma+\gamma^2$.

The FOIL method can be used to expand the product of two sums or differences. FOIL stands for first, outer, inner, last, and represents the products one must calculate.

F: Multiply the first terms: $(\underline{2\pi}+7)(\underline{4}+3\pi) = 2\pi \times 4 = 8\pi$.

O: Multiply the outer terms: $(\underline{2\pi}+7)(4+\underline{3\pi}) = 2\pi \times 3\pi = 6\pi^2$.

I: Multiply the inner terms: $(2\pi+\underline{7})(\underline{4}+3\pi) = 4 \times 7 = 28$.

L: Multiply the last terms: $(2\pi+\underline{7})(4+\underline{3\pi}) = 7 \times 3\pi = 21\pi$.

Add terms to get $8\pi+6\pi^2+28+21\pi$.

Finally, group like terms to get $6\pi^2+29\pi+28$.

To test yourself, factor the final expression and show it yields the simplified expression with which we started. This is one way to check your work for any careless mistakes.

2.2.4 Solving Equations

Solving an equation involves isolating a variable on one side (by convention, the left side of the equals sign) and all other variables and constants on the other side. One does so by performing the same calculations on both sides of the equation such that one ends up isolating the variable of interest. This often takes multiple steps and there is almost always more than one way to arrive at the solution. As an example, the equation $y = 2x$ is already solved for y. If we want to solve that equation for x, we need to do some algebra. Start with

$$y = 2x.$$

Divide both sides of the equation by 2, yielding

$$\frac{y}{2} = x.$$

Rewrite the equation:

$$x = \frac{y}{2}.$$

Note that we can go about this in a more convoluted fashion:

$$y = 2x.$$

Divide both sides by x, yielding

$$\frac{y}{x} = 2.$$

Divide both sides by y:

$$\frac{1}{x} = \frac{2}{y}.$$

Multiply both sides by x:

$$1 = x(\frac{2}{y}).$$

Multiply both sides by $\frac{y}{2}$:

$$\frac{y}{2} = x.$$

Now rewrite:

$$x = \frac{y}{2}.$$

That is hardly efficient, but the good news is that we ended up at the same place, though we would have had to be careful that neither x nor y was equal to zero when dividing by them. We also got some practice in manipulating an equation. Here are a few useful techniques for those rusty in their algebra.

1. **Focus on the variable of interest**. Work on isolating the variable you care about, and don't worry so much about what this does to the rest of the equation.

2. **Combine all like terms**. Simplifying equations is easiest when you sort out all the noise and add together like terms.

3. **Check your answer**. Substitute the value that you obtain into the original equation to make sure that your answer is correct and that you didn't make a careless mistake.

4. **Make use of identities**. Remember $\frac{a}{b} \times \frac{b}{a} = 1$ and $a - a = 0$. That means you can multiply by the first and add the second at all times, whenever convenient, without changing the equation.

5. **Operate on both sides in the same manner**. Adding the same number to each side or multiplying each side by the same number won't change the equation.

2.2.4.1 Solving Quadratics

Solving quadratic polynomials requires learning how to complete the square and/or knowing the quadratic equation.

Completing the Square

Many quadratic equations that you will face can be solved relatively easily by **completing the square**. The basic intuition for solving these is to isolate the variable and its square and then add a value to each side of the equation to "complete the square."

To see what we are trying to accomplish, it helps to begin with a simple example. Sometimes we are presented with a quadratic equation that factors into a squared term, i.e., $(x - n)^2 \pm c$, where c is some constant. Consider the quadratic $x^2 - 6x + 5$. We can factor this into $(x - 3)^2 - 4$ (use the FOIL method to verify). We can then rewrite this equation as $(x - 3)^2 = 4$. Finally, by taking the square root of both sides we, can solve for x:

$$\begin{gathered}(x-3)^2 = 4 \Rightarrow \\ x - 3 = \pm 2 \Rightarrow \\ x = 5 \text{ or } x = 1.\end{gathered}$$

Note that this quadratic equation will have two solutions in the real numbers, or zero, but not one.[6] In other words, the cardinality of the solution set for a quadratic equation will be zero or two. An example of a quadratic equation with no real solutions (i.e., no solutions in the real numbers) is $x^2 + 1 = 0$.[7]

Solving a quadratic by factoring it into a squared term $\pm$ a constant and then taking the square root is quick, but most quadratics cannot be factored in integers so easily. However, we can transform any quadratic using the following steps to "complete the square" (i.e., transform it into a squared term $\pm$ a constant) and then solve for x by taking the square roots.

[6] Quadratic equations with real coefficients will always have two solutions in complex numbers; if these solutions are complex they will come in pairs, i.e., $a \pm bi$.

[7] The solutions of this equation are $\pm i$.

1. Start with a quadratic in your variable of interest (we'll say it's x) and move the constant to the right-hand side. Divide through by the coefficient on x^2. So if you have $2x^2 - 4x - 2 = 0$, you get $x^2 - 2x = 1$.

2. Divide the coefficient on x by 2 and then square it. Add that value to both sides of the equation. So now you have $x^2 - 2x + 1 = 1 + 1$.

3. Factor the left-hand side into a "(x $\pm$ some term) squared" form and simplify the right-hand side. So now you have $(x - 1)^2 = 2$.

4. Take the square root of both sides (remember that when you take the square root of a number, the solution is always $\pm$, because the square of a negative number is a positive number). So now you have $x - 1 = \pm\sqrt{2}$.

5. Solve for x. So the solutions are $x = 1 + \sqrt{2}$ and $x = 1 - \sqrt{2}$.[8]

Let's work another example. Consider the quadratic

$$x^2 + 8x + 6 = 0.$$

The first thing you might try is factoring to see if it yields the "(x $\pm$ some term) squared" form. This quadratic does not, so we turn to completing the square. The first step is to isolate the squared term and variable by subtracting 6 from each side (note the coefficient on x^2 is already 1):

$$x^2 + 8x = -6.$$

Next we need to add the square of half of the value in front of x to both sides. The value in front of x is 8, so we divide 8 by 2 and then square the result $4^2 = 16$. Thus, to complete the square we need to add 16 to each side:

$$x^2 + 8x + 16 = 10.$$

We then perform step 3 and factor the left-hand side of the equation:

$$(x + 4)^2 = 10.$$

We now need to take the square root of each side, which gives us

$$x + 4 = \pm\sqrt{10}.$$

Now we can solve for x by subtracting 4 from each side. Our final answer is

$$x = -4 + \sqrt{10} \text{ and } x = -4 - \sqrt{10}.$$

[8] Irrational solutions such as $1 \pm \sqrt{2}$ will also always come in pairs.

The Quadratic Formula and Equation

Completing the square is one method for solving a quadratic equation, but it is not the only one, and you will sometimes encounter equations that are rather complicated to solve by completing the square. For example, if you are faced with $x^2 + \sqrt{15}x - 1 = 0$, you will not want to calculate half of $\sqrt{15}$, and then square it, add it to both sides, and try to factor the result. Instead, you will want to turn to the quadratic equation and formula.[9]

During your high school algebra courses you were probably required to memorize the quadratic equation and formula. The general form of a quadratic equation is[10]

$$ax^2 + bx + c = 0.$$

The general solutions to this equation are

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

These solutions are called the quadratic formula. The formula can be derived from the equation by completing the square (we ask you to do this below, in an exercise). When we obtain values for x, we call these values the roots of the equation. For our purposes, this formula is used when completing the square is made difficult by fractions, decimals, or large numbers. We refer to a and b as the coefficients and c as the constant.

What are the roots of the quadratic equation

$$1.4x^2 + 3.7x + 1.1 = 0?$$

To find the solutions, we first list the values for a, b, and c:

$$a = 1.4, b = 3.7, c = 1.1.$$

Then we plug the values for a, b, and c into the quadratic formula:

$$x = \frac{-3.7 \pm \sqrt{3.7^2 - 4 \times 1.4 \times 1.1}}{2.8}.$$

Using a calculator to solve, we find that $x = -.341$ or $x = -2.301$.

These problems can be cumbersome because it is somewhat more difficult to check them. Two pieces of advice, though, can help you minimize mistakes. First, make sure to follow order of operations (PEMDAS, remember?). Second, you can go online and use a quadratic equation solver and plug in the values for a, b, and c to verify the accuracy of your computations. We discuss this further in Section 3.

[9] It is possible to use either method—completing the square or the quadratic equation and formula—to solve a quadratic equation. But one is more likely to make errors using the latter than the former, and so many people find completing the square preferable as long as they are not faced with an unusual (e.g., radical) value in front of the x term.

[10] The equation is true when $a \neq 0$. When $a = 0$ and $b \neq 0$, it is a simple linear equation with solution $x = -\frac{c}{b}$. If both a and b are zero, the equation is false unless $c = 0$ as well.

2.2.5 Inequalities

To solve inequalities, we have to discuss a few extra properties.

First, all pairs of real numbers have exactly one of the following relations: $x = y$, $x > y$, or $x < y$.

Adding any number to each side of these relations will not change them; this includes the inequalities. That is, inequalities have the same addition and subtraction properties as equalities such that if $x > y$, then $x + a > y + a$ and $x - a > y - a$.

The properties for multiplication and division for inequalities are a bit different than for equalities. For multiplication, if a is positive and $x > y$, then $ax > ay$. If a is negative and $x > y$, then $ax < ay$. For division, if a is positive and $x > y$, then $\frac{x}{a} > \frac{y}{a}$. If a is negative and $x > y$, then $\frac{x}{a} < \frac{y}{a}$. Multiplying or dividing an inequality by zero is not allowed.

To summarize, this means that you flip the $<$ or $>$ sign when multiplying or dividing by a negative.

Example: Solve for y: $-4y > 2x + 12$. First, we want to isolate y by itself on the left side of the equation. We divide both sides by -4, which gives us $y < -\frac{x}{2} - 3$. Dividing by a -4 flips the $>$ sign to $<$. If we do not know the value of x, then we can leave it in this form.

2.2.6 Review: Avoiding Common Errors

We have included a list of some common mistakes people make when solving equations as a sort of help file for when you are struggling to find the right answer. Below this list, we've included some websites you may go to for extra examples or help.

That said, please remember that the World Wide Web is dynamic, and the links below will become dated. We found them using search engines, and you will be able to do the same.

Sign errors: Sign errors are probably the most common mistakes. Most people think of this $(-)$ sign as a negative sign. This is part of the problem. Tackling math as if it were a foreign language is the best way to approach learning the fundamentals of mathematics. This sign $(-)$ is best thought of as "the opposite of," or, in other words, "the additive inverse." (Recall that every integer and real number has an additive inverse that, when added to the number, produces zero.) When reading an equation such as $-x + y = 7$, you should say in your head, "the opposite of x added to y is 7." The reason for thinking of this sign as "the opposite of" is twofold. First, it can help you find mistakes in your work. Second, it will help you deal with situations such as $-x = 7$. You will easily interpret this as the opposite of x equals 7, so x must be the additive inverse of 7, which is -7.[11]

[11] If this is a bit confusing, remember that the number line has 0 in the middle, positive

Only changing one side or term in an equation: Think of an equation as a scale or seesaw. Whatever you do to one side you must do to the other.[12] An equation must be in equilibrium. If you divide one side by 12, you must divide the other by 12. In addition, you must divide all terms on both sides by 12.

Not distributing: Always distribute across addition (and subtraction). If you have an expression such as $4x(2 + 6y + 3t)$, many people simply multiply $4x$ by 2. Each term inside a parentheses must be multiplied by what is outside the parentheses. The correct expression is $8x + 24xy + 12xt$.

Distributing with radicals and exponents: Radicals and exponents have different rules, which we discuss in depth in the next chapter. They do not follow the same rules as multiplication and addition. For example, $\sqrt{9 + 16}$ is not the same as $\sqrt{9} + \sqrt{16}$. Also, as discussed above, $(\alpha + \beta)^2$ is not the same as $\alpha^2 + \beta^2$.

You can find other lists of common errors at several websites. For example, Eric Schecter maintains a page of the most common math errors by undergraduates (`http://atlas.math.vanderbilt.edu/~schectex/commerrs/`). See Schecter's page for entries on "multiplying by a negative one and other sign errors," "loss of invisible parentheses," "everything is additive," and "everything is commutative."

Other common algebra mistakes include canceling terms instead of factors, misunderstanding fractions, and misunderstanding negative and fractional components. See `http://tutorial.math.lamar.edu/pdf/algebra_Cheat_Sheet.pdf`.

Beyond this, some of you may be interested in more practice, especially with algebra. One of the authors finds Huettenmueller (2010) a useful resource, but there are a number of other self-teaching guides. You can also find a number of useful resources available on the Web. We recommend `http://www.purplemath.com/`, `http://mathworld.wolfram.com/`, and `http://math.com/`. Wikipedia (`http://en.wikipedia.org/`) also has many good entries for mathematical concepts, though many of these can substantially be found elsewhere.

For more information on set theory, see Peter Suber's "A Crash Course in the Mathematics of Infinite Sets" (`http://www.earlham.edu/~peters/writing/infapp.htm`). Oregon State's "Field Guide to Functions" (`http://oregonstate.edu/instruct/mth251/cq/FieldGuide/`) is a good guide to functions (the topic of the next chapter). R.H.B. Exell's page on relations is also useful for

integers falling to the right of 0 and negative integers falling to the left. Any number has an opposite on the number line that is equidistant from zero. So the opposite of 8 is -8 and the opposite of -9 is 9. Thinking of negatives in these terms will also help you deal with absolute values.

[12] Of course, as noted above, adding zero to one side or multiplying one side by one is acceptable, as these are identities and leave the value of the expression unchanged. This may be may be useful when working with fractions.

a more detailed introduction (http://www.jgsee.kmutt.ac.th/exell/Logic/Logic42.htm).

2.2.7 Why Should I Care?

algebra is the set of rules one uses to manipulate equations that have variables in place of numerical values, whereas arithmetic is the set of rules we use to manipulate equations made of numerical values. arithmetic is thus essential for making specific calculations, but algebra is needed if we want to study general concepts. You care about algebra for the same reason you care about arithmetic: people use it to communicate their ideas precisely, and they often assume you can do algebraic operations in your head. To follow along, then, you need to do the algebra. This is true in both the study of statistics and the study of formal theory. If you do not master this basic algebra, you will get lost. Solving equations and simplifying inequalities in order to find the range of solutions also proves highly useful in both game theory and statistics. Indeed, as we explain in Chapter 12, which introduces vector algebra and matrix algebra, the algebra covered here (which is called scalar algebra) is a foundation for both vector and matrix algebra.

2.3 COMPUTATIONAL AIDS

Throughout this book we assume that you will be performing all mathematical manipulations by hand, or at most using a (simple) calculator. We believe this is pedagogically important: one needs to be able to do the relevant calculations oneself in order to understand them; if one doesn't understand them, then one doesn't really know what one is saying; and if one doesn't know what one is saying, there is very little point to formalizing one's concepts with mathematics at all. So this book is intended to lead you through doing the calculations yourself. That said, it is often helpful to have access to computational aids for arithmetic, algebra, and the later topics in this book. One reason simply is as a check for your work. We all make mistakes, and it is nice to have a second pair of eyes, so to speak, to check one's work. A second reason is to increase speed once one is sure of one's understanding. As the techniques of math become more familiar to you, the boundary of your skills will expand, and you will want to devote more of your time to the harder stuff rather than simple Algebraic manipulation. Computational aids can help with this. Finally, a third reason is to help with the writeup. Some aids allow output in formats that may be easily converted to word processors or typographical languages such as LaTeX (http://www.latex-project.org/).

There are many computational aids out there. Some are freeware, meaning you can download them from the Internet, or use them in your browser directly, with no further obligation. Some examples of these include Eigenmath (http://eigenmath.sourceforge.net/) and Maxima (http://maxima.sourceforge.net/), along with the website http://www.wolframalpha.com/, which allows you to input expressions directly into your browser. There are also some useful tools at http://www.math.com/students/tools.html, including a function plotter. LaTeXis a free typographical language that is very good at typesetting math; we wrote this book using it, and thus were able to deliver it typeset to Princeton University Press, retaining greater control over its "look and feel" and reducing production costs. Various options exist to make LaTeXmore user-friendly, e.g., LyX (http://www.lyx.org/).

Other tools are potentially more powerful, but they are also more expensive. However, they can have more functionality in some areas. If you are located at a university with access to them, they can be worthwhile to try. Mathematica (http://www.wolfram.com/mathematica/) and Maple (http://www.maplesoft.com/products/maple/) work well with symbolic math, and Matlab (http://www.mathworks.com/products/matlab/) is well suited to matrix algebra.

2.4 EXERCISES

2.4.1 Arithmetic Rules

Complete the following equations:

1. $x^1 =$ ______.
2. $-a \times (-b)^2 =$ ______.
3. $\sum_{i=1}^{4} x_i =$ ______.
4. $\prod_{m=6}^{9} x_m =$ ______.
5. $4! =$ ______.
6. $z^4 =$ ______.
7. $\sqrt[2]{9} =$ ______.
8. $\sqrt[3]{27} =$ ______.
9. $\left(\frac{3(2-4)}{2+3}\right)^3 =$ ______.

2.4.2 Ratios, Proportions, Percentages

10. Represent the following as a ratio, a proportion, and a percentage:

    a) Latinos relative to all others: African American 98,642; Asian 62,346; Caucasian 436,756; Latino 105,342; Other 32,654.

    b) Independent registered voters relative to Republicans: Democrat 432; Independent 221; Republican 312.

c) Republican relative to Democrat from no. 2.

11. If the Latino population shrunk to 100,322 in no. 1 above, what would be the percentage change in the Latino population?

12. If the other populations remained constant, what would be the percentage change in the proportion of Latinos to all others?

13. If voter turnout in the United States in 1996 was 56% and in 2000 it was 62%, what was the percentage change in turnout from 1996 to 2000?

14. Express these two quantities as a simplified ratio: 18 and 12.

2.4.3 Algebra Practice

15. Simplify into one term the following expressions:

    a) $xz + yz$.
    b) $mn + ln - pn$.
    c) $z \times y \times x - 2 \times y \times x$.
    d) $(z + x) \times y \times \frac{1}{z}$.

16. Simplify this expression as much as possible: $\frac{2x^2+20x+50}{2x^2-50}$.

17. Simplify this expression: $\frac{5+17x+4x+7}{42x}$.

18. Add these fractions: $\frac{2g+13}{3g} + \frac{4g-5}{4g}$.

19. Factor: $-7\theta^2 + 21\theta - 14$.

20. FOIL: $(2x - 3)(5x + 7)$.

21. Factor: $q^2 - 10q + 9$.

22. Factor and reduce: $\frac{\beta-\alpha}{\alpha^2-\beta^2}$.

23. Solve: $15\delta + 45 - 6\delta = 36$.

24. Solve: $.30\Omega + .05 = .25$.

25. Solve: $11 = (y + 1)2 + (6y - 12y)\frac{7}{2y}$.

26. Solve: $-4x^2 + 64 = 8x - 32$.

27. Complete the square and solve for x: $x^2 + 14x - 14 = 0$.

28. Complete the square and solve for y: $\frac{1}{3}y^2 + \frac{2}{3}y - 16 = 0$.

29. Solve using the quadratic formula: $2x^2 + 5x - 7$.

30. Derive the quadratic formula by completing the square for the equation $ax^2 + bx + c = 0$.

31. Solve: $-\delta > \frac{\delta+4}{7}$,

Chapter Three

Functions, Relations, and Utility

Hagle (1995, p. 7) opines that "functions are valuable to social scientists because most relationships can be modelled in the form of a function." We would add that functions are valuable for those political scientists who want to make *specific* theoretical claims and/or use statistics to test the implications of theories of politics. In other words, functions are valuable because they are explicit: they make very specific arguments about relationships. In addition, functions play a key role in developing statistical models.

What is a function? Functions may be defined in several ways, each developed more fully below. To get us started, functions provide a specific description of the association or relationship between two (or among several) concepts (in theoretical work) or variables (in empirical work). In other words, a function describes the relationship between ordered pairs (or n-tuples) arising from sets under special conditions (specified below).

That said, some students can come away from an introduction to relations and functions with a misguided notion that the key to developing sound theory is to master a wide array of functions and then see which one applies to a given theoretical or statistical problem. One might characterize this as a "toolbox" approach to political science, where different functions are hammers and wrenches to be tried here and there until one finds one that works. Perhaps such an approach would yield insight, but we are not sanguine: one's thinking about politics is unlikely to be usefully informed simply by mastery of different functional forms. Instead, a general working knowledge of functions can be used to sharpen one's thinking and bring greater specificity to one's theories of politics. In particular, learning to be able to translate verbal conjectures into graphs and/or equations that represent those conjectures is a valuable skill to develop. That skill is essential for anyone who wants to do formal modeling. Finally, such a working knowledge is critical to mastering the material in statistics courses and will help one select appropriate statistical models for hypothesis testing.

The first section discusses functions in general and elaborates on some of their properties. The second illustrates various functions of one variable; most of these can be readily generalized to multiple variables. The third section covers properties of relations in the milieu in which they are most typically seen in political science—individual preferences—and introduces the utility representation that underlies all of game theory. This serves as another example of the use of functions in political science (an empirical example appears at the end of the second section) and provides us with an opportunity to mention correspondences briefly as well. Readers with stronger math backgrounds should be able to skim the first two sections, but may not have seen the material in the third before.

3.1 FUNCTIONS

Recall from Chapter 1 that relations allow one to compare variables and expressions (or concepts). This is a general idea, but some relations are considerably more specific about the comparison. In particular, any relation that has a unique value in its range (we'll call these y values) for each value in its domain (we'll call these x values) is a function. Put differently, all functions are relations, but only some relations are functions. Another way to put this is that functions are subsets of relations. That said, political scientists do not often distinguish between relations and functions, and the term "function" is often used loosely to cover both relations and functions. Alternatively, you may encounter relations described as "set functions" and functions (as defined here) described as "point functions." More precisely, a relation that assigns one element of the range to each element of the domain is a **function**, while one that assigns a subset of the range to each element of the domain is a **correspondence**. We will focus largely on functions here, as they are the most commonly used by political scientists. However, correspondences are commonly used in game theory, and we discuss them briefly in Section 4.

More formally, a function maps the values measuring one characteristic of an object onto values measuring another characteristic of the object. Stated in set theoretic terms, a function is a relation such that (1) for all x in A, there exists a y such that (x, y) is an ordered pair in the function, and (2) if (x, y) and (x, z) are in the function, then $(y = z)$. In other words, if the value x is mapped to the value y by a function, and the value x is also mapped to the value z by the same function, then it follows that y and z are the same value. If $y \neq z$, then it is not a function but a correspondence.

Note that some equations with which you are familiar from middle school and high school math are either functions or correspondences. We review some examples below.

One can use both equations and graphs to describe functions. If you can develop an ability to translate your verbal conjectures into functions, you will have sharper, more explicit conjectures. Thus, developing the ability to work comfortably with both equations and their graphs will prove very valuable for developing your own theories about politics.

3.1.1 Equations

The linear equation $y = a + bx$ is the best-known and most frequently used function in political science.[1] We discuss it below. Here we want to remind you of the manner in which functions can be represented using equations. One often encounters equations of the form $x^2 + y^2 = 1$ or $\frac{y}{x} = 3$. We can use the rules covered previously to isolate y on the left-hand side (LHS),[2] yielding $y^2 = 1 - x^2$ and $y = 3x$. It turns out that the first of these equations is not a function while the second is, and we demonstrate that below where we introduce graphs.

You will hopefully recall the notation $y = f(x)$, which is read "y is a function of x." This is **implicit** notation that simply states that values of x are associated with singular values of y. Here we call x the **argument** of the function. But we do not know what the specific function is, so if we were given the values of x we could not produce the values of y. An **explicit** function describes the mapping of values in the domain to values in the range. For example, if we were given the explicit function $y = 3x$, then we could map the values of y for any given set of x values. In empirical work we typically refer to the x here as the independent or exogenous variable and the y as the dependent or endogenous variable, as it depends on and is affected by x.

3.1.2 Graphs

As noted above, we can graph relations and functions. If we plot the values of a set (or concept or variable) on the horizontal axis and the values of another set that shares ordered pairs with the first set on the vertical axis, then we can plot the intersection of each pair's values with a point in the space defined by the axes. Such a graph is known as a Cartesian, or xy, graph and is quite common. You will recall such graphs from arithmetic and algebra courses. The horizontal axis is also referred to as the x-axis (or domain) and the vertical axis is also known as the y-axis (or range).

The graph of the relation $x^2 + y^2 = 1$ forms a circle through the values 1, -1 on both axes, as depicted in Figure 3.1.[3] Note that this is not a function: all values in the domain (x) produce two different values in the range (y). If this were not true, it would not form a circle. If you do not find this apparent, select a value of x and plot the value for y in Figure 3.1.

Now consider the equation $y = 3x$, shown in Figure 3.2. The graph of this equation is a straight line moving through the origin and up to the right. No matter what x values we plug into the equation we get a unique value of y. As such, the equation is a function. Note that we can make use of a graph to determine whether an equation is or is not a function: if we can draw a vertical

---

[1] It may surprise you that though it is often referred to as a linear function, the linear equation is not a linear function, as strictly defined in mathematics. We discuss this below.

[2] Subtract x^2 from both sides in the first case and multiply both sides by x in the second.

[3] Because it makes a circle with a one-unit radius, it is known as the unit circle.

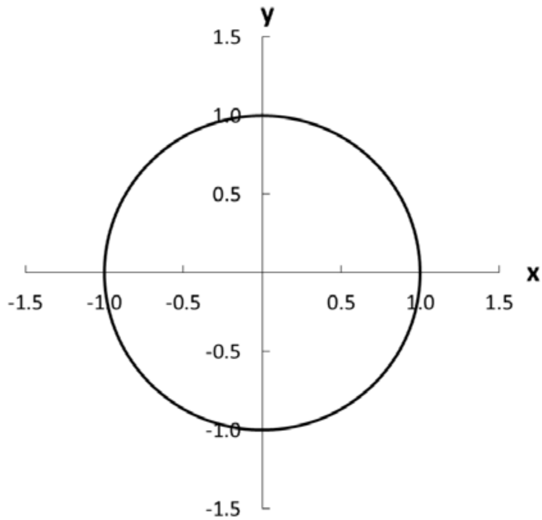

Figure 3.1: Graph of the Unit Circle

line at *any* point on the graph that intersects the curve at more than one point, then the equation is not a function.

3.1.3 Some Properties of Functions

As we go on, several properties of functions will be important. We cover inverse and identity functions, monotonic functions, and functions in more than one dimension, saving continuity for Chapter 4 and function maxima and minima, along with concave and convex functions, for Part II of the book. To begin, we expand our notation for a function slightly. We define the function f as $f(x) : A \to B$. This is often read as "f maps A into B." You've already seen the first part, which just means that the variable x is an input to the function $f(x)$, which spits out some value. Sometimes we assign this value to a variable y, as in $y = f(x)$, and sometimes we just leave it as $f(x)$, where it is understood that the function $f(x)$ may itself be a variable or a constant. For example, $f(x) = 3x$ is a variable, whereas $f(x) = 3$ is a constant.

The A and the B in the function's definition are new, but not conceptually. A here is the **domain** of the function, that is, the set of elements over which the function is defined. In other words, we draw our values of x from this set, and the function needs to produce a value for each element of x in this set. The most common domain political scientists use is the real numbers, $\mathbb{R}$, but there are numerous other domains you will see. B is known as the **codomain**, and it specifies the set from which values of $f(x)$ may be drawn. Depending on A

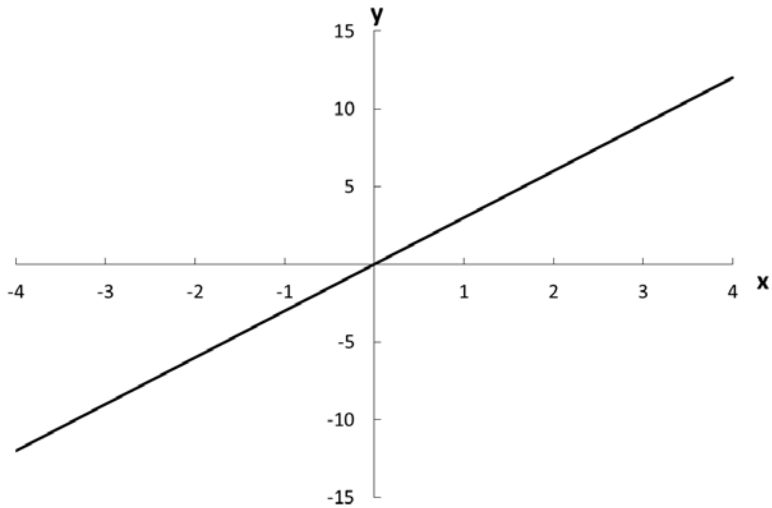

Figure 3.2: Graph of $y = 3x$

and f, though, not all the values in B may be reached. The set of all values actually reached by running each $x \in A$ through f is known as the **image**, or **range**, and it is necessarily a subset of B.

This may be confusing, so let's consider an example. Let $f(x) = x$. This function maps x to itself, and so does really nothing.[4] If $A = \mathbb{R}$, then $B = \mathbb{R}$, and the codomain and the range (or image) are exactly the same, since every real number is just mapped to itself. Now instead keep $B = \mathbb{R}$, indicating that the function f is real-valued, but let $A = (0, 1)$, or the set of all real numbers between zero and one, exclusive. In this case the image (or range) is just $(0, 1)$, which is the only part of B reached by the function, given the domain A.

One can chain multiple functions; this is called **function composition**. This is written either as $g \circ f(x)$ or $g(f(x))$ and is read as "g composed with f" or more commonly g of f of x. If we have $f(x)A \to B$ and $g(x) : B \to C$, then the full definition is $g \circ f(x) : A \to C$. Composition of functions is associative $(f \circ (g \circ h) = (f \circ g) \circ h)$, but not always commutative ($f \circ g$ does not always equal $g \circ f$). One takes a function composition in stages: first one computes $f(x)$ for each x to get a set of y, and then one takes $g(y)$ for each of these y. For more than two functions that are composed, first plug each x into the innermost function, then plug the output of this into the next innermost function, and so on until you've finished with all the functions. For example, if $f(x) = 2x$ and $g(x) = x^3$, then $g \circ f(x) = (2x)^3 = 8x^3$, whereas $f \circ g(x) = 2(x^3) = 2x^3$.

[4]This function is called the identity function, and we return to it below.

Table 3.1: Identity and Inverse Function Terms

Term	Meaning
Identity function	Elements in domain are mapped to identical elements in codomain
Inverse function	Function that when composed with original function returns identity function
Surjective (onto)	Every value in codomain produced by value in domain
Injective (one-to-one)	Each value in range comes from only one value in domain
Bijective (invertible)	Both surjective and injective; function has an inverse

3.1.3.1 Identity and Inverse Functions

Why does this all matter? To answer that, we need a couple more definitions as we need to introduce identity and inverse functions, as well as some other terms. Table 3.1 summarizes those terms.

A function is **surjective** or **onto** if every value in the codomain is produced by some value in the domain.[5] Our first example was surjective, because every point in $\mathbb{R}$ was reached by some point in the domain (the same point, in the example). The second was not surjective, as nothing outside $(0, 1)$ in the codomain was reached.

A function is **injective** or **one-to-one** if each value in the range comes from only one value in the domain.[6] We already knew that each $x \in A$ produced only one $f(x)$; otherwise it wouldn't be a function. This tells us that this property goes both ways: each $y \in f(x)$ comes from only one $x \in A$. Both of our examples for the identity function are injective; the function is just a straight line. In contrast, $f(x) = x^2$ would not be injective on the same domain as, for example, $y = 4$ is the result of plugging both $x = 2$ and $x = -2$ into the function (it would be injective if we confined ourselves to real numbers no less than zero, though).

If a function is both injective and surjective (one-to-one and onto), then it is **bijective**. A bijective function is **invertible**, and so has an inverse. This inverse is the payoff of our definitions, as it allows us to take a y and reverse our function to retrieve the original x. How do we do this? First we (re)define an **identity function**: $f(x) = x, f(x) : A \to A$, where we have made the domain and codomain identical, as we saw in our earlier example. This function merely returns what is put into it and is just like multiplying each element in our domain by one (or adding zero to each element), hence the use of the word identity.

[5] Formally, it is surjective if $\forall b \in B, \exists a \in A \ni f(a) = b$ (for all b in B there exists an a in A such that the function of a is b).

[6] Formally, $\forall a, c \in A, \forall b \in B$, if $f(a) = b$ and $f(c) = b$, then $a = c$.

The **inverse function** is the function that when composed with the original function returns the identity function. That is, it undoes whatever the function does, leaving you with the original variable again. The inverse is $f^{-1}(x) : B \to A$, and remember to be *very* careful not to confuse it with $(f(x))^{-1} = \frac{1}{f(x)}$. Thus, in symbols, the inverse is defined as the function $f^{-1}(x)$ such that $f^{-1} \circ f(x) = x$, or just $f^{-1}(f(x)) = x$. The inverse does commute with its opposite $f(f^{-1}(x)) = f^{-1}(f(x))$. For example, if $f(x) = 2x + 3$, a bijective mapping, then its inverse is $f^{-1}(x) = \frac{x-3}{2}$. We can check this both ways: $f^{-1}(f(x)) = \frac{(2x+3)-3}{2} = \frac{2x}{2} = x$ and $f(f^{-1}(x)) = 2\left(\frac{x-3}{2}\right) + 3 = x - 3 + 3 = x$.

3.1.3.2 Monotonic Functions

Some functions increase over some subset of their domains as x increases within this subset. Others decrease over the same subset, and the rest increase over some x and decrease over others, depending on the value of x. If a function never decreases and increases for at least one value of x on some set $C \subseteq A$, it is an **increasing function** of x on C, while if it never increases and decreases for at least one value of x on some set $C \subseteq A$, it is a **decreasing function** of x on C. If a function increases always as x increases within C it is a **strictly increasing function** on C; if it decreases always as x increases within C it is a **strictly decreasing function** on C. Strictly increasing and strictly decreasing functions are injective. We sometimes call a function that does not decrease (but may or may not increase ever) a **weakly increasing function**, and a function that does not increase (but may or may not decrease ever) a **weakly decreasing function**.

You will sometimes encounter the term **monotonic function** in statements such as "y increases monotonically as a function of x." Monotonicity is the characteristic of order preservation—it preserves the order of elements from the domain in the range. A monotonic function is one in which the explained variable either raises or retains its value as the explanatory variable(s) rises. Thus it is an increasing function across its entire domain. A strictly monotonic function is strictly increasing over its entire domain. Table 3.2 summarizes these concepts.

We provide several examples of monotonic functions in the next section. All affine and linear functions with positive coefficients on x are strictly monotonic, as are exponential functions, logarithms, cubic equations, etc. Ordered sets can also be monotonic or strictly monotonic. An example of two ordered sets with a monotonic, but not a strictly monotonic, relationship is $\{1, 2, 3, 4, 5\}$, $\{10, 23, 23, 46, 89\}$. Monotonic functions have many nice properties that will become apparent as you study both statistics and game theory.

3.1.3.3 Functions in More Than One Variable, and Interaction (Product) Terms

Thus far we have (primarily) simplified things by focusing on the idea that y was a function of one variable. Unfortunately, few (if any!) political relationships

Table 3.2: Monotonic Function Terms

Term	Meaning
Increasing	Function increases on subset of domain
Decreasing	Function decreases on subset of domain
Strictly increasing	Function always increases on subset of domain
Strictly decreasing	Function always decreases on subset of domain
Weakly increasing	Function does not decrease on subset of domain
Weakly decreasing	Function does not increase on subset of domain
(Strict) monotonicity	Order preservation; function (strictly) increasing over domain

are so simple that they can be described usefully as a function of one variable. As such, we need to be able to use functions of two or more variables, such as $y = f(x_1, x_2, x_3)$ or $z = f(x, y)$.

Graphs of the function of one variable are straightforward, and graphs of the function of two variables are feasible (though many of us begin to struggle once we have to start thinking in three dimensions). Consider two variables multiplied by one another, also known as a product term. Product terms are a commonly used nonlinear function. Consider the plot of $y = 3xz$, in Figure 3.3, and observe that it produces a plane with a changing slope rather than a plane with a constant slope.

Another way of saying the same thing is that the relationship of x on y is different (stronger or weaker) depending on the value of z. Further, the strength of the impact of z on y also depends on the value of x. That is what is meant by interaction: x and z interact with one another to produce y. You will learn in your statistics course how to properly specify statistical models to test interaction hypotheses.[7]

As another example, consider the three-dimensional plot of the linear function $y = 3x + z$, depicted in Figure 3.4.

Graphs of the function of three or more variables, however, become terribly complex and generally are not used, though there are some exceptions. Instead of using graphs, analysis of multiple variable functions focuses on equations.

Luckily, the specification of equations in more than one variable is not much more complicated than that in one variable. You've already seen some of the notation, e.g., $f(x, y)$. The rest just accounts for the more complex domain

[7] We note here that one would not want to estimate the model $y = \alpha + \beta(xz) + e$ because doing so would produce a biased estimate of β. Rather, one would want to include the variables x and z in the model as well (Blalock Jr., 1965; Friedrich, 1982; Braumoeller, 2004; Brambor, Clark, and Golder, 2006). This will be discussed in your statistics courses.

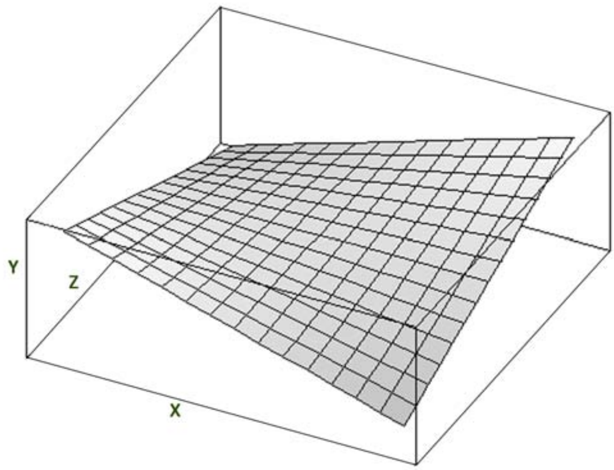

Figure 3.3: Graph of $y = 3xz$

that is present when there is more than one variable. If there are n variables, denoted x_1 through x_n, and the set from which each variable is drawn is called A_1 through A_n, respectively, then the domain of the function is the Cartesian product $A_1 \times A_2 \times \ldots \times A_n$. The formal definition of the function is $f(x_1, \ldots, x_n) : A_1 \times \ldots \times A_n \to B$. To get any value of f you just plug in the values of all the input variables. Most of the concepts discussed above are either directly applicable or have analogues in the multidimensional case, though there is more complexity involved. For example, properties such as continuity can be defined for each input variable independently. We save discussion of the properties of multi-dimensional functions most relevant to us until Part V of the book, however.

3.1.4 Why Should I Care?

A basic understanding of functions is critical to any political scientist who wants to be able to make specific causal conjectures. Making specific causal conjectures is useful because it increases one's ability to evaluate whether relevant evidence is at odds with one's theory (i.e., improves hypothesis testing; Popper, 1959, pp. 121–23) and it facilitates communication with other scholars (Cohen and Nagel, 1934, pp. 117–20). Vagueness is antithetical to science, and stating hypotheses as functions helps one eliminate vagueness. Further, statistical inference is a powerful tool for hypothesis testing, and functions are one of the building blocks on which statistics is constructed. Finally, game theory makes extensive use of functional forms to represent preferences and payoffs, as we'll see in Section

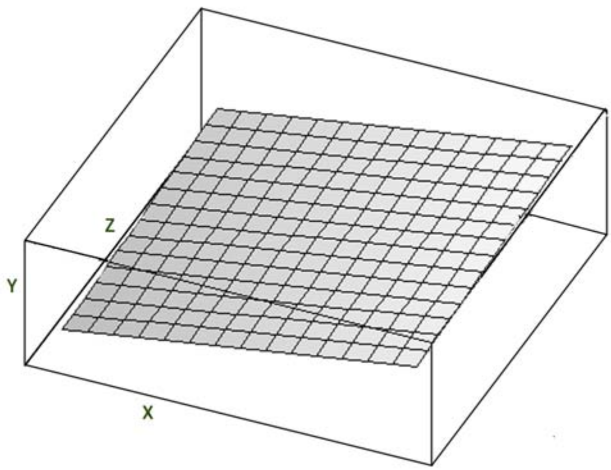

Figure 3.4: Graph of $y = 3x + z$

3 of this chapter. For these reasons, the properties of functions we discuss in this section are fundamental, as they have substantive meaning in the settings in which we are using the functions. Monotonocity in one's preferences, for example, means that someone always prefers more to less. This is very different from having what is known as an ideal point, in which case moving away from the ideal in *either* direction is not preferred.

3.2 EXAMPLES OF FUNCTIONS OF ONE VARIABLE

Political scientists are generally interested in the relationships among multiple variables. Nevertheless, in this section we begin with associations where y is a function of one x. These functions extend readily to more than one variable, as noted above.

3.2.1 The Linear Equation (Affine Function)

You encountered the additive linear equation back in algebra classes: $y = a+bx$. Technically, this is an **affine function**, though it is frequently referred to as a linear function. We discuss the technical distinction between the two below. For now, let's review some basics.

In the equation $y = a + bx$, a and b are constants.[8] The constant a is the

[8]Recall that in this equation, y is a function of only one variable, x. Therefore a and b cannot be variables and must be constants.

intercept, or in terms of the graph, where the function crosses the vertical (y) axis (i.e., the value of y when $x = 0$). The constant b is the slope of the line, or the amount that y changes given a one-unit increase in x. That is, a one-unit increase in x produces a $1b$-unit increase in y, a three-unit increase in x produces a $3b$-unit increase in y, etc.

One might conjecture that the probability that an eligible voter casts a ballot in a US presidential election is a linear function of education.[9] Let p_v represent the probability of voting and ed represent education level: $p_v = a+b(ed)$. In this function a represents the likelihood that someone without any formal education turns out to vote, and b indicates the impact of education on the probability of voting. Shaffer (1981, p. 82) estimates a model somewhat like this, and we can borrow his findings for illustrative purposes, yielding $p_v = 1.215 + 0.134 \times ed$. The intercept of 1.215 makes little sense,[10] but we ignore that for this example. Shaffer's education measure has four categories: 0–8 years of education, 9–11 years, 12 years, and more than 12 years. A slope (i.e., b) of 0.134 suggests that as we move from one category to another (e.g., from 0–8 years to 9–11 years, or from 12 years to more than 12 years), the probability that someone votes rises by 0.134. So if this linear model and its results are accurate, the typical adult with a college education has roughly a 0.4 greater probability of voting in a US presidential election than the typical adult without any high school education.[11]

The linear equation states that the size of the impact of x on y is constant across all values of x. For example, in the above example the impact of x on y is roughly 0.13. Since the relationship is linear, that means that a shift from 0–8 years of education to 9–11 years of education increases the probability of voting in a national election by $\sim$.13, and a shift from 9–11 years to 12 years also produces an increase of $\sim$.13, as do shifts from 12 years of education to more than 12 years of education. Nonlinear functions, which we discuss in the third subsection, specify that the size of the impact of x on y varies across values of x.

[9] We recognize that people are very unlikely to posit such a claim. We offer it not as a reasonable conjecture but simply as an illustration.

[10] A value of 1.2 is nonsense because the intercept represents the probability of voting when a person has had zero education. Since probabilities by definition have a range from zero to one, any probability above one is nonsense. This is but one reason that this may be an unrealistic example. In your statistics courses you will learn a number of reasons why this estimate is nonsense.

[11] You will learn how to do these sorts of calculations in your statistics courses. For those who want a brief description, you need to calculate two values and then determine the distance between them. More specifically, multiply the slope (0.134) by the first value in the comparison, someone with no high school education: $1 \times 0.134 = 0.134$. Now multiply the slope by the second value in the comparison, someone with a college education: $4 \times 0.134 = 0.536$. Finally, take the difference of these two probabilities of voting (i.e., subtract the probability that a citizen without a high school education votes from the probability that a college educated citizen votes) to get $0.536 - 0.134 = 0.402 \sim 0.4$.

3.2.2 Linear Functions

Mathematicians make distinctions that few political scientists employ. We review them for the purpose of helping you avoid confusion when you read "mathematically correct" presentations. In particular, we distinguish between affine functions (discussed above), linear equations, and linear functions (discussed here). As suggested above, a **linear equation** is an equation that contains only terms of order x^1 and $x^0 = 1$.[12] In other words, only x and 1, multiplied by constants, may appear on the right-hand side (RHS) of a linear equation. This means that the RHS of a linear equation is an affine function. Linear *functions* are not affine functions; e.g., they do not permit a translation (the x^0 term).

The formal definition of a **linear function** is any function with the following properties:

- Additivity (aka superposition): $f(x_1 + x_2) = f(x_1) + f(x_2)$,
- Scaling (aka homogeneity): $f(ax) = af(x)$ for all a.

Additivity states that the impact of a sum of variables is equivalent to the sum of the impacts of those variables. The scaling property, on the other hand, states that the size of the input is proportional to the size of the output.

Let's begin by comparing the linear function $y = \beta x$ with the affine function $y = \alpha + \beta x$ along these criteria. The additivity property states that $f(x_1 + x_2) = f(x_1) + f(x_2)$. So we substitute the RHS of each $y = \ldots$ equation for the parts in the parentheses (i.e., $f(\cdot)$) and see if that statement is true. If it is, the property is met. We begin with the linear function $y = f(x) = \beta x$. To determine whether it meets the additivity property, we need to replace x with $x_1 + x_2$, following the additivity property equation above, and determine whether the equality is true:

$$\begin{aligned} f(x_1 + x_2) = \beta(x_1 + x_2) &= \beta x_1 + \beta x_2, \\ \beta x_1 + \beta x_2 &= f(x_1) + f(x_2). \end{aligned}$$

As one can see, the equality is true. Now we'll try the linear equation (or affine function), under the assumption that $\alpha \neq 0$: $y = f(x) = \alpha + \beta x$. Again, we replace x with x_1 and x_2, in accord with the additive property equation, and see whether the equality is true:

$$\begin{aligned} f(x_1 + x_2) = \alpha + \beta(x_1 + x_2) &= \alpha + \beta x_1 + \beta x_2, \\ f(x_1) + f(x_2) &= (\beta x_1 + \alpha) + (\beta x_2 + \alpha), \\ \alpha + \beta x_1 + \beta x_2 &\neq 2\alpha + \beta x_1 + \beta x_2. \end{aligned}$$

It is not true; the RHS and LHS differ by α. So the linear equation (or affine function) does not have the additive property, but the linear function does.

[12]Order refers to the highest exponent in the polynomial.

Now let's consider the scaling property, which states that $f(ax) = af(x)$. Let's begin with the linear function $y = f(x) = \beta x$:

$$\begin{aligned} f(ax) = \beta(ax) &= a\beta x, \\ a\beta x &= af(x). \end{aligned}$$

So, the linear function satisfies the scaling property. What about the linear equation (i.e., affine function)?

$$\begin{aligned} f(ax) = \alpha + (\beta(ax)) &= \alpha + a\beta x, \\ af(x) &= a\alpha + a\beta x, \\ \alpha + a\beta x &\neq a\alpha + a\beta x. \end{aligned}$$

This property doesn't hold either because $\alpha \neq a\alpha$. Again, the linear function satisfies the property, but the affine function (linear equation) does not.

The only difference between the two functions is the constant, α. Recall that α represents the value where the function crosses the vertical (y) axis. If it crosses at zero, then the two functions are equivalent. Thus, a linear function must cross the vertical axis at the origin (i.e., where x and y have a value of zero). You might recall that ratio level measurement requires a meaningful zero value, whereas interval level measurement does not, and that division and multiplication operations are valid on ratio level measures but not on interval level measures. Linear transformations require preservation of the order of the variables, the scale, and the zero, and only linear functions meet such criteria. Affine transformations preserve order and scale, but not the placement of zero.

Above we noted that political (and other social) scientists frequently refer to the linear equation $y = \alpha + \beta x$ as a linear function. *Technically*, this is inaccurate, but it is a rather fine mathematical point. The linear equation does produce a line, and a linear transformation with the affine function preserves order and scale, with the exception of the intercept. And that is all most political scientists are typically trying to indicate when they talk about linear functions and linear transformations. That said, there are some applications (e.g., time series analysis) where the proper definition of a linear function is important, and we raise the discussion here so as not to later confuse those who go on to study those issues in more detail.

3.2.3 Nonlinear Functions: Exponents, Logarithms, and Radicals

Technically speaking, nonlinear functions are all those that do not meet the two properties we just discussed. Practically speaking, nonlinear functions are all those that are neither linear nor affine: those functions that describe (the graph of) a curve that is not a line. For example, $y = \cos(x)$, in Figure 3.5, is a nonlinear function. Functions with exponent terms, including quadratics and other polynomials, are the most commonly used nonlinear functions in political science. Logarithms are another commonly used class of nonlinear functions, as are roots (or radicals). We briefly introduce the relationship among these functions and then turn our attention to graphing these functions and using them in algebra.

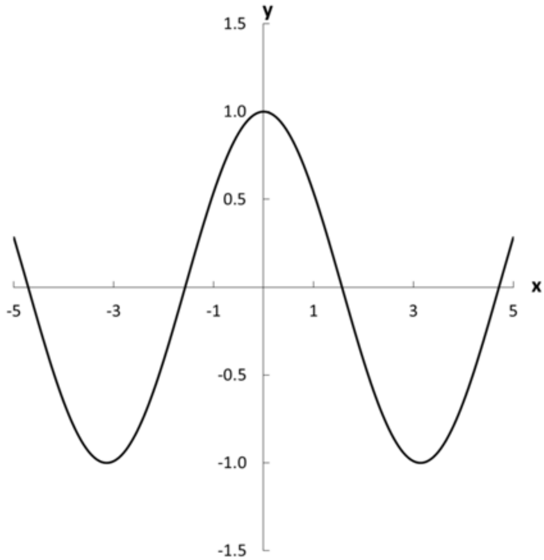

Figure 3.5: Graph of $y = \cos(x)$

Exponents, logarithms, and roots are related: one can transform any one such function into a representation of one of the others. In fact, in high school you may have focused on doing that. More specifically, when two of the following variables in the equation $b^n = x$ are known, one can solve for the unknown using

- Exponents to solve for x,
- Logarithms to solve for n,
- Radicals to solve for b.

That said, we will not focus on the relationship among the functions as political scientists do not frequently make use of those relationships.[13] Instead, we introduce each function and its notation, discuss their graphs, and then describe algebraic manipulations.

3.2.3.1 Exponents and the Exponential Function

As notation, **exponents** (aka power functions) are a shorthand for expressing the multiplication of a number by itself: $x^3 = x \times x \times x$. More generally, $x^n = x \times$

[13]Those interested in studying this might find the following Wikipedia entries useful: `http://en.wikipedia.org/wiki/Logarithm`, `http://en.wikipedia.org/wiki/Radical_(mathematics)#Mathematics`, `http://en.wikipedia.org/wiki/Exponential_function`.

$x \times x \ldots x$ (n times). This is all familiar, but you may be less familiar with other exponential notation: $x^{-n} = \frac{1}{x^n}, x^{\frac{1}{n}} = \sqrt[n]{x}$. In words, x to a negative power represents the fraction "1 divided by x^n" and x raised to a fraction represents a root of x, where the root is determined by the value in the denominator of the exponent. Perhaps an easier way to remember this is that a negative exponent indicates that one *divides* (rather than multiplies) the term by that many factors. Similarly, a fractional exponent indicates that one takes the n^{th} *root* rather than multiplying the term n times. Mixed exponents work similarly. So $x^{\frac{2}{3}} = \sqrt[3]{x^2}$ and $x^{\frac{-3}{2}} = \frac{1}{\sqrt[2]{x^3}}$. Finally, $x^0 = 1$.[14]

Nonlinear functions with exponents are of interest to political scientists when we suspect that a variable x has an impact on y, but that the strength of the impact is different for different values of x. The best way to see this is to look at the graphs of some functions with exponents.

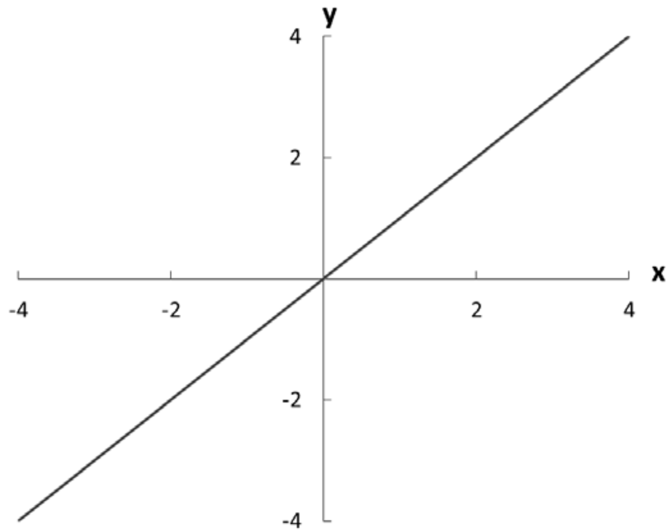

Figure 3.6: Graph of $y = x$

Consider the graphs of the functions $y = x$ and $y = x^2$, in Figures 3.6 and 3.7. The linear function produces a line with a constant slope: if we calculate the change in y due to a one-unit change in x, it does not matter what point on the x-axis we select; the change in y is the same.[15] However, the slope of the curve for $y = x^2$ is not constant: the impact of x on y changes as we move along the x-axis (i.e., consider different values of x). To be more concrete, a one-unit

[14]This holds for all $x \neq 0$, but people often treat $0^0 = 1$ as if it were true when they are simplifying equations.

[15]Another way to make this point is to observe that linear functions meet the scaling property.

increase from 0 to 1 produces a one-unit increase in y, but a one-unit increase from 2 to 3 produces a five-unit increase in y, and a one-unit increase from 5 to 6 produces an 11-unit increase in y. Thus, the impact of x on y increases over the range of x.

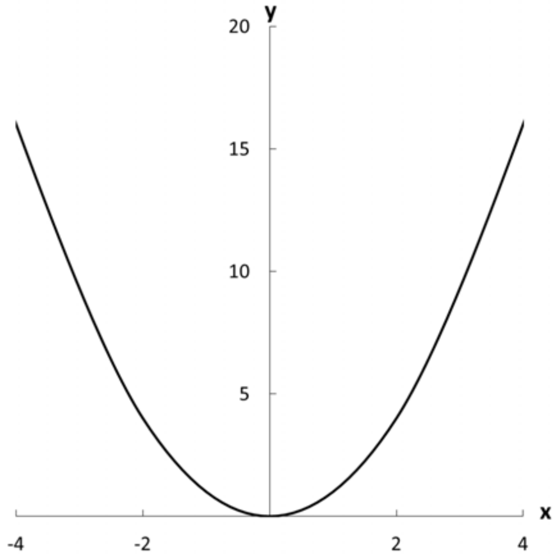

Figure 3.7: Graph of $y = x^2$

This has important implications for developing theory. If reflection, deduction, or inspiration leads one to conjecture that a causal relationship between two concepts is constant over the range of values for the causal concept, then a linear or affine relation represents that conjecture. However, if one suspects that the strength of the relationship varies across the values of the causal concept, then a nonlinear relation is needed. As we discuss below, exponential terms play an important role in quadratic and other polynomial functions.

We covered some of these above, but below is a list of the algebraic rules that govern the manipulation of exponents.

Multiplication: to calculate the product of two terms with the same base one takes the sum of the two exponents:

$$x^m \times x^n = x^{m+n}.$$

To see that this is so, set $m = 3$ and $n = 4$ and write it out:

$$x^3 \times x^4 = (x \cdot x \cdot x) \times (x \cdot x \cdot x \cdot x) = x \cdot x \cdot x \cdot x \cdot x \cdot x \cdot x = x^7.$$

This works when m and n are positive, negative, or zero. When the bases are different, you can simplify the expression *only when the exponents are the same.*

In this case, multiplication is distributive:

$$x^m \times z^m = (xz)^m.$$

To see why, set $n = 2$, and note that $x^2 \times z^2 = x \cdot x \cdot z \cdot z = (x \cdot z) \times (x \cdot z) = (xz)^2$.[16]

Last, when both the base and the exponent are different, you cannot simplify to a single term. Thus, e.g.,

$$x^m \times z^n \neq (xz)^{m+n}.$$

Assume that $m = 2$ and $n = 3$, and write the expressions out to see that this is so:

$$x^2 \times z^3 = (x \times x) \times (z \times z \cdot z) = z((x \times z) \times (x \times z)) \neq (xz)^5.$$

One cannot combine the terms fully.[17] To return to the point made above, if we assume that $m = n = 3$, then when we write it out we get:

$$x^3 \times z^3 = (x \cdot x \cdot x) \times (z \cdot z \cdot z) = (x \cdot z) \times (x \cdot z) \times (x \cdot z) = (xz)^3.$$

To determine the **power of a power**, one multiplies the exponents. For example,

$$(x^m)^n = x^{mn}.$$

To see that this is so, let's assign $m = 2$ and $n = 3$, and write out:

$$(x^2)^3 = x^2 \times x^2 \times x^2 = x \cdot x \cdot x \cdot x \cdot x \cdot x = x^6.$$

Division: to calculate the quotient of two terms with the same base and different powers, one takes the difference of the exponents:

$$\frac{x^m}{x^n} = x^{m-n}.$$

To see why this is so, recall that

$$\frac{1}{x^n} = x^{-n}.$$

We can therefore write out:

$$\frac{x^m}{x^n} = x^m x^{-n} = x^{m-n}.$$

[16]Note that this assumes that multiplication is commutative; hence this will not hold for matrix multiplication, as we'll see in Part IV of the book.

[17]If this illustration is not clear to you, then assign values to x and z (say, 2 and 3) and work it out. It will become clear that one can take the product when the exponents are equal and the bases are different, but one cannot take the product when both the exponents and bases are different. Note that we can simplify to a degree: $x^2 \times z^3 = (xz)^2 z$, but this is not usually helpful.

We can assign the values $m = 2$ and $n = 3$ and verify

$$\frac{x^2}{x^3} = \frac{x \cdot x}{x \cdot x \cdot x} = \frac{1}{x}$$

and

$$\frac{x^2}{x^3} = x^2 x^{-3} = x^{-1} = \frac{1}{x}.$$

When the bases are different, one can simplify only if the exponents are the same. When the exponents are the same, one raises the fraction to that power:

$$\frac{x^m}{z^m} = \left(\frac{x}{z}\right)^m.$$

Put differently, like multiplication, division is distributive when the bases are different and the exponents are the same.[18]

Recall that $x^0 = 1$. We can now demonstrate this by observing that $\frac{x^n}{x^n} = 1$. Observe that $\frac{x^n}{x^n} = x^{n-n} = x^0$. Since anything divided by itself equals one, it follows that $x^0 = 1$ (except when $x = 0$).

This covers x^a, but what about a^x? This is called an **exponential**. The one most commonly used sets $a = e$, where e is the base of the natural logarithm, or $e \approx 2.7183$ (to four decimal places). This is the **exponential function**, written as $y = \exp(x)$ or $y = e^x$.[19] We discuss the base of the natural logarithm, and its relation to the exponential function, below; in Figure 3.8 we graph the exponential function.

3.2.3.2 Quadratic Functions

Quadratic functions are nonlinear functions that describe a parabola. More specifically, if y is a quadratic function of x, then $y = \alpha + \beta_1 x + \beta_2 x^2$. In other words, quadratic functions describe a relationship where a variable (y) is a function of the sum of another variable (x) and its square (x^2).[20]

---

[18] In fact, other than having to remember not to divide by zero, multiplication and division have basically the same properties. The same is true for addition and subtraction. Thus one need remember only the properties of multiplication and addition.

[19] Both notations are common, and they are equivalent expressions.

[20] It is worth observing that many political scientists refer to a quadratic function as linear. For example, in a regression course you may encounter the claim that $y = \alpha + \beta_1 x + \beta_2 x^2 + e$ is a linear model. That is true: it is a linear *model*. When discussing regression models people frequently distinguish between models that are *linear in parameters* from those that are *linear in variables*. A regression model that contains a quadratic function (e.g., $y = \alpha + \beta_1 x + \beta_2 x^2 + e$) is linear in parameters but nonlinear in variables. Put differently, if we plot the relationship between x and y, the plot will be nonlinear: it is not linear in variables. But the parameters of the quadratic function have the properties of an affine function (to see this, set $z = x^2$ and rewrite the linear model as $y = \alpha + \beta_1 x + \beta_2 z$), and if we assume that $\alpha = 0$, then they have the properties of a linear function. Returning to models, the model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + e$ is linear in parameters and variables (as long as we assume that x_1 and x_2 are not nonlinear transformations of one another (e.g., $x_1 \neq x_2^n$)).

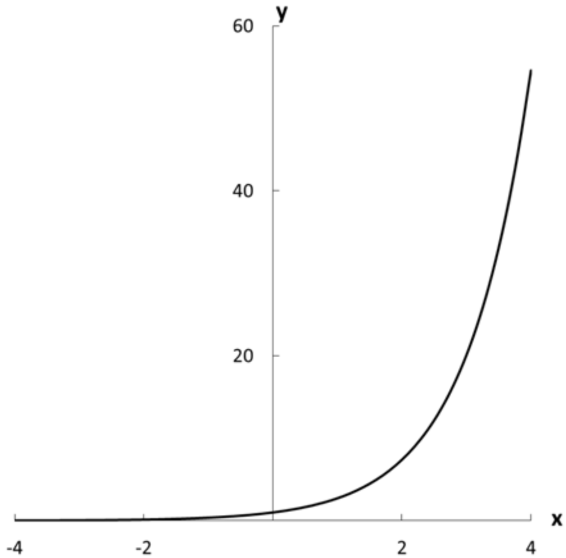

Figure 3.8: Graph of $y = e^x$

Note that since y is a function of only one variable, x, we can graph the function in two dimensions. If we set $\beta_2 < 0$, then we get a curve shaped like an inverse U (i.e., a concave parabola) as depicted in Figure 3.9. Switching the sign of β_2 produces a U-shaped curve (i.e., a convex parabola).

What sort of theoretical expectations might one want to sharpen by stating them as a quadratic relationship? Speaking generally, a quadratic function is quite useful for depicting relationships where we think the impact of an independent variable is positive (negative) for low values of the independent variable, flat for middle-range values, and negative (positive) for high values. Put differently, when one thinks that there is some (often unknown) threshold at which the relationship between two concepts (variables) switches (i.e., from positive to negative or from negative to positive), one might consider whether the quadratic can represent our conjecture.

For example, many scholars have hypothesized that rebellion will be low in countries that exert little to no government coercion *and* in countries that exhibit high levels of government coercion. Where will one find rebellion? This conjecture suggests that it will be highest among those countries that engage in mid-range levels of coercion (e.g., Muller and Seligson, 1987). If we let r represent rebellion and c represent coercion, then this conjecture can be represented as follows: $r = \alpha + \beta_1 c - \beta_2 c^2$.

Another example is the conjecture that the extent to which governments are transparent (i.e., noncorrupt) varies nonlinearly with the level of political competition. More specifically, over the range from authoritarian to democratic

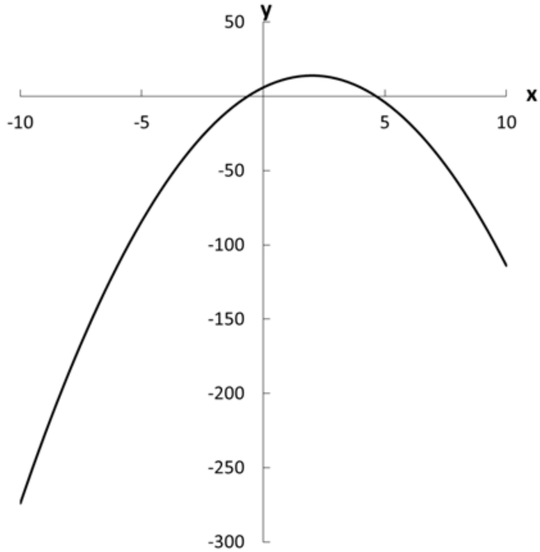

Figure 3.9: Graph of $y = 6 + 8x - 2x^2$

polities, transparency (e.g., the absence of bribery) is relatively common at both endpoints and least common in mixed polities that have a mix of autocratic and democratic institutions (e.g., Montinola and Jackman, 2002). If we allow t to stand for transparency and p for polity type, then we can represent that conjecture with the following quadratic equation: $t = \alpha + \beta_1 p + \beta_2 p^2$.

Finally, note that if we invert the concept we are trying to explain (i.e., flip the scaling of the dependent variable), we can represent the argument by flipping the signs on the quadratic (x^2) term. Thus, if we reconceptualize rebellion as quiescence, q, then we can write $q = \alpha + \beta_1 c + \beta_2 c^2$, and if we reconceptualize transparency as corruption, k, then we can write $k = \alpha + \beta_1 p - \beta_2 p^2$.

3.2.3.3 Higher-Order Polynomial Functions

Polynomial functions have the following general form: $y = \alpha + \beta_1 x + \beta_2 x^2 + \ldots + \beta_n x^n$, where n is an integer less than infinity. So both linear and quadratic functions are polynomials. Higher-order polynomials are those possessing powers of x greater than the reference. In this case, we are referring to the presence of cubed and higher terms. Like quadratics, higher-order polynomials are nonlinear: they describe curves, such as the cubic polynomial in Figure 3.10. More specifically, one can use them to explicitly represent the expectation that there are two or more thresholds over which the relationship between two concepts (variables) changes.

With the exception of the quadratic, polynomial functions are not very common in political science, though Mukherjee (2003) and Carter and Signorino

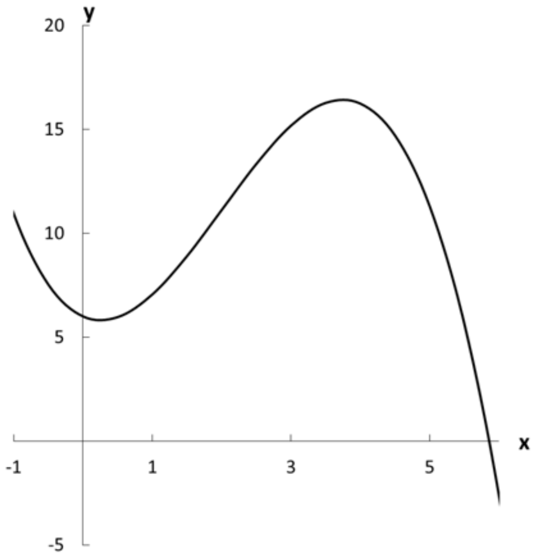

Figure 3.10: Graph of Cubic Polynomial

(2010) are exceptions. Mukherjee studies the relationship between the size of the majority party and central government expenditures in parliamentary democracies. A majority party is one that has at least 50% of the seats in the legislature and thus can govern without having to form a coalition with other parties. The basic underlying idea is that there are two different thresholds at work between the number of seats the majority party holds in the legislature and the size of government spending. First, as the number of seats held by the majority party rises from a bare majority (i.e., 51% of the legislature), spending declines, because it takes more and more legislators to defect and bring down the government.

Yet, while Mukherjee expects an initial negative relationship as the size of the majority party increases above a bare majority, he expects the relationship to quickly become positive (perhaps at around 56% of the legislative seats). Expenditures rise because the party has greater electoral safety and thus can take greater risks of alienating other parties' constituents by more greatly rewarding its own constituents. Yet he does not argue that this incentive to spend more remains as party size grows beyond the supermajority threshold (roughly 67% of the seats).

Instead, Mukherjee expects the relationship between majority party size and government expenditures to again turn negative (above the supermajority threshold) because the size of the population that the majority can tax without suffering electorally shrinks. That is, as majority party size rises beyond the supermajority threshold, the number of constituents that support other parties grows sufficiently small that it becomes increasingly difficult to write legislation that transfers income from those people to one's own constituents. He uses a cubic polynomial, $GovExp = \alpha - \beta_1(SizeMajParty) + \beta_2(SizeMajParty)^2 - \beta_3(SizeMajParty)^3$, to represent his verbal argument, and the results of his empirical analysis are consistent with his conjecture.

Carter and Signorino (2010) propose the use of a cubic polynomial to model time dependence in binary pooled cross-sectional time series data. Though it sounds complex, it is a fairly straightforward proposal. One takes the measure of time in one's data (perhaps the year) and, like Mukherjee, includes the three-termed polynomial in the regression equation. They show that if the dependent variable can take only two values (e.g., absence or presence of war) and the researcher has both cross-sectional data (e.g., all the countries in the world) measured over time (e.g., 1816–2005), then the cubic polynomial of time will control for what is called "temporal dependence" in the regression model.

More generally, then, polynomial functions are appealing because one can use them to make specific claims about threshold effects. That is, when theorizing leads one to expect that the relationship between two variables changes across the values of one of the variables, then a polynomial function might help one make a more specific (and more easily testable and falsifiable) claim.

3.2.3.4 Logarithms

Logarithms can be understood as the inverses of exponents (and vice versa). They can be used to transform an exponential function to a linear one, or a linear function to a nonlinear one in which the impact of one variable on another declines as the first variable rises in value. The logarithm (or log) tells you how many times to multiply its base a in order to get x, where a is a positive real number not equal to 1. If we denote the log with base a by $\log_a x$, then we have $a^{\log_a x} = x$ and $\log_a a^x = x$. Similarly, we can see that if $\log_a x = b$, then $a^{\log_a x} = a^b$, since the exponents are the same, and thus $x = a^b$. This lets us transition between logs and exponents readily.

Logs can be written in any base, though the most common are base 10 and the natural log. The base for the natural log is the $e \approx 2.7183$ from the exponential function. The concept of the base of a number is abstract and often confuses students. This owes in part to the commonality of the base 10 system in our lives. It is, after all, how we write numbers: we use 0 through 9, and at 10, 100, 1,000, and so on, we add a digit. You may also be familiar with binary from living in an age of computers, however. In binary one uses only 0 and 1, and at 2, 4, 8, and further powers of 2, one adds a digit. The base of a log is just an extension of this idea. We won't go into why e is one of the most common bases of logs used, though you are free to explore that topic on your own, of course. Rather, we'll just note that the natural log is *usually* identified with the notation ln, and log base 10 is *generally* denoted log, though some people use log to denote the natural logarithm. Throughout this book ln indicates natural log and log denotes log base 10.

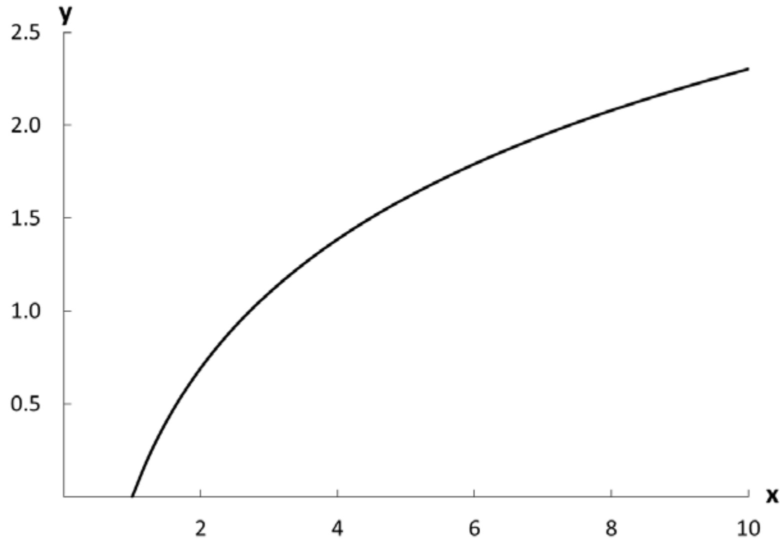

Figure 3.11: Graph of $y = \ln(x)$

Let's look at graphs of $y = \ln(x)$ and $y = \log(x)$ in Figures 3.11 and 3.12. Note that the impact of x on y diminishes as x increases, but it never becomes zero, and it never becomes negative. Theoretically, the log functions are very appealing precisely because of this property.[21] If you suspect, for example, that education increases the probability of voting in national elections, but that each additional year of education has a smaller impact on the probability of voting than the preceding year's, then the log functions are good candidates to represent that conjecture. Why? If p_v is "probability of voting" and ed is "years of education," then $p_v = \alpha + \beta ed$ specifies a linear relationship where an additional year of education has the same impact on the probability of voting regardless of how many years of education one has had. By contrast, $p_v = \alpha + \beta ed^2$ represents the claim that the impact of education on the probability of voting rises the more educated one becomes. Neither of these functional forms captures the verbal conjecture. But if we take the log of an integer variable such as "years of education," we transform the relationship between p_v and ed from a linear one to a nonlinear one where the impact of an additional year of education declines the more educated one becomes: $p_v = \alpha + \beta(\ln(ed))$.

There are several algebraic rules for logs that are important to know.[22] First,

[21] This is called *concavity*, and we will discuss it more in Part II of this book.
[22] The following holds for logarithms of any base, not just the natural log.

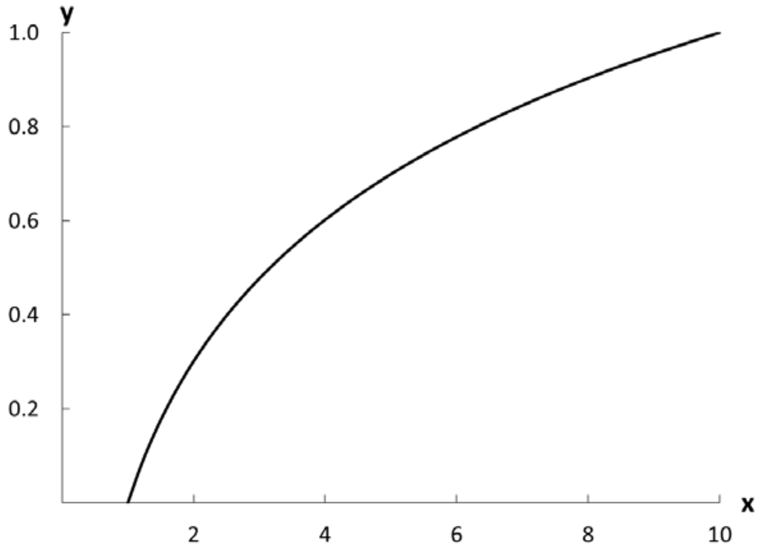

Figure 3.12: Graph of $y = \log(x)$

note that the log is not defined for numbers less than or equal to zero.[23] Further, $\ln(1) = 0$ (i.e., $\ln(x) = 0$ when $x = 1$), and $\ln(x) < 0$ when $0 < x < 1$.

Second, the log of a product is equal to the sum of the logs of each term, and the log of a ratio (or fraction) is the difference of the logs of each term:

$$\ln(x_1 \cdot x_2) = \ln(x_1) + \ln(x_2), \text{ for } x_1, x_2 > 0$$

and

$$\ln \frac{x_1}{x_2} = \ln(x_1) - \ln(x_2), \text{ for } x_1, x_2 > 0.$$

Note that addition and subtraction of logs *do not* distribute:

$$\ln(x_1 + x_2) \neq \ln(x_1) + \ln(x_2) \text{ ,for } x_1, x_2 > 0,$$

and

$$\ln(x_1 - x_2) \neq \ln(x_1) - \ln(x_2), \text{ for } x_1, x_2 > 0.$$

These equations cannot be simplified further. Thus, if one takes the log of both sides of the equation $y = \alpha + \beta_1 x_1 + \beta_2 x_2$, the solution is *not* $\log y = \log \alpha + \log \beta_1 + \log x_1 + \log \beta_2 + \log x_2$ but $\log y = \log(\alpha + \beta_1 x_1 + \beta_2 x_2)$.

[23]This follows from the identity $a^{\log_a x} = x$. Assume $a > 0$, and that $x \leq 0$. Let $\log_a x = b$. Then we have $a^b \leq 0$ for $a > 0$, which is impossible, implying that b is undefined. Thus the log is defined only for $x > 0$. Other properties can also be derived from this identity and the rules on exponents we stated earlier.

Third, the log of a variable raised to a power is equal to the product of the exponent value and the log of the variable:

$$\ln(x^b) = b\ln(x), \text{ for } x > 0.$$

Finally, as $x > 0$ approaches 0 (so x is small), the log of $1+x$ is approximately equal to x:[24]

$$\ln(1 + x) \approx x, \text{ for } x > 0 \text{ and } x \approx 0.$$

Political scientists generally use log functions to represent conjectures that anticipate a declining impact of some x over some y as x increases in value. For example, Powell (1981) studies the impact of electoral party systems on mass violence (as well as other forms of system performance). In the study, he controls for both the population size and per capita gross national product (GNP). The basic ideas are that (1) countries with larger populations will produce more riots and deaths from civil strife and (2) those with greater economic output per person will produce fewer riots and deaths from political violence. But Powell (and most social scientists) do not expect these relationships to be linear: an increase in population from 1,000,000 people to 2,000,000 people will have a greater impact on riots and deaths than will an increase in population from 100,000,000 to 101,000,000. Similarly, an increase from \$500 to \$1,500 GNP per capita is expected to have a greater impact on the number of deaths and riots a country will typically experience than an increase from \$18,000 to \$19,000 GNP per capita. That is, Powell hypothesizes that the positive and negative effects of population and economic output, respectively, on civil strife will decline as the value of population and economic output rises.[25] We can thus write Powell's expectations as: $CS = X + \ln(P) - \ln(G)$, where CS represents civil strife, X represents the party system variables that Powell considers, and P and G represent the control variables population and per capita GNP, respectively.[26] While a log function is only one of many one could use to convert those verbal claims to a specific mathematical statement, it is a common function that has often performed well in statistical tests.

Wallerstein (1989) provides another example. He explores the determinants of cross-national difference in labor unionization rates. One of the variables Wallerstein expects to have an effect is the size of the potential union membership (i.e., labor force). If we let U indicate unionization rate, L the size of the labor force, and X the other variables that he considers, we can represent his expectation as $U = \ln(L) + X$.[27] Why expect a nonlinear log relationship? Wallerstein explains that "using the log of the potential membership implies that the percentage increase, rather than the absolute increase, matters for

[24]This follows from a Taylor expansion of the log. We discuss this in Part II of this book.

[25]In other words, the marginal effect of these variables is decreasing. We discuss marginal effects at length in Part II of this book.

[26]Readers familiar with regression analysis in statistics might expect a representation like this: $CS = \alpha + \beta_1 X + \beta_2 \ln(P) - \beta_3 \ln(G) + \epsilon$.

[27]Using a regression representation, the argument is $U = \alpha + \beta_1 \ln(L) + \beta_2 X + \epsilon$.

union density" (p. 490). This argument stems from the equation for the difference in logs. If $\Delta \ln(L) = \ln(L_t) - \ln(L_{t-1})$ is the change in the natural log of the labor force variable, L, then $\Delta \ln(L) = \frac{\ln(L_t)}{\ln(L_{t-1})}$. This is a ratio rather than a difference in different values of the labor force.

The two most common usages of the log function are (1) to model the nonlinear expectation that the size of the effect of one variable on another declines as the second variable rises in value and (2) to model the expectation that the relative increase of a variable over time has a linear impact on another variable.

3.2.3.5 Radicals (or Roots)

Roots (sometimes called **radicals**) are those numbers represented by the radical symbol: $\sqrt[n]{\ }$. They are (almost) the inverse functions of x raised to the power n: $\sqrt[n]{x^n} = x = (\sqrt[n]{x})^n$ as long as n is odd or $x \geq= 0$.[28] Functions with radicals are nonlinear: $y = \sqrt[n]{x}$. Some roots are integers: $\sqrt[2]{9} = 3$. However, most are not: $\sqrt[2]{3} \approx 1.732050808$. Figure 3.13 graphs the function for $n = 3$ over the range $x = [1, 4]$.

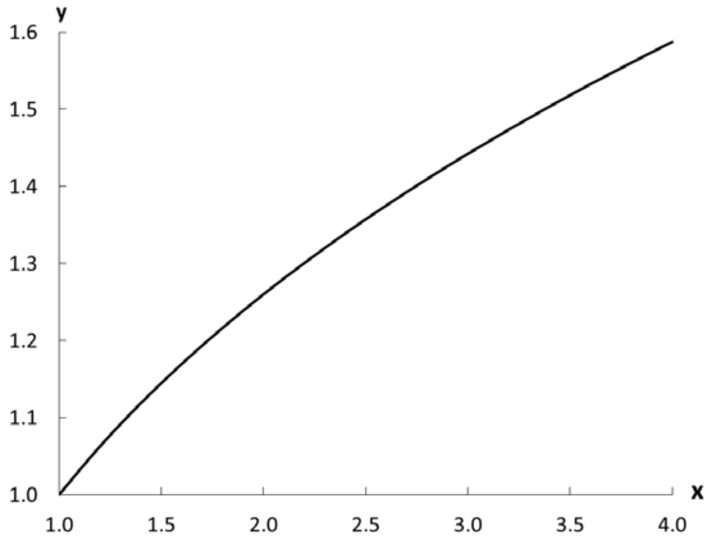

Figure 3.13: Graph of $y = x^{\frac{1}{3}}$

As noted above, radicals can also be expressed as fractional exponents: $\sqrt{x} = x^{\frac{1}{2}}$. We can express this more generally by observing that $\sqrt[n]{x^p} = (\sqrt[n]{x})^p = x^{\frac{p}{n}}$. When $n = 2$, we typically do not write the 2 in $\sqrt{x^p}$.

[28]Even roots (i.e., n is even) are undefined for negative values of x in the real numbers. They are defined in the complex number system using the definition of the imaginary number i, where $i = \sqrt[2]{-1}$.

Although roots do not play a large role in political science, one encounters them from time to time. For example, Gelman, Katz, and Bafumi (2004) explore a common assumption in the literature on the fairness (with respect to representation) of weighted voting systems such as the US Senate, where people living in states with smaller populations (e.g., Maine) have a greater influence on the votes cast in the Senate than people living in states with larger populations (e.g., Illinois). The conventional assumption is that all votes are equally likely (i.e., that voting is random), and a common indicator used to measure the "voting power" of an individual citizen is the Banzhaf index: $\frac{1}{\sqrt{N}}$. Gelman, et al. argue that this index "(and, more generally, the square-root rule) overestimates the probability of close elections in large jurisdictions" (p. 657). As an alternative indicator they recommend the fraction $\frac{1}{N}$.

To do algebra with roots one needs to memorize the following rules.

Addition and Subtraction

One cannot in general add or subtract two radicals. So:

$$\sqrt[n]{x} + \sqrt[n]{x} \neq \sqrt[n]{x+x} \text{ for } n > 1.$$

For example, $\sqrt{2} + \sqrt{2} = 2\sqrt{2} > 2 = \sqrt{4} = \sqrt{2+2}$.

Note that one cannot sum the roots, either:

$$\sqrt[n]{x} + \sqrt[n]{x} \neq \sqrt[n+n]{x+x} \text{ for } n > 1.$$

Observe that $\sqrt{9} + \sqrt{9} = 3 + 3 = 6 \neq \sqrt[4]{18}$ because $6^4 \neq 18$.

This is also so when the variables and roots are different, e.g.,

$$\sqrt[a]{x} + \sqrt[b]{y} \neq \sqrt[a+b]{x+y} \text{ for } a, b > 1.$$

To see this, note that $\sqrt[2]{9} + \sqrt[3]{8} = 3 + 2 = 5 \neq \sqrt[5]{17}$ because $5^5 \neq 17$.

The only exception is when one side would be zero, either because at least one of x or y is zero or because we are using subtraction and $x = y$.

Multiplication and Division

One can determine the product of two radicals *only* when they have the same order. In such a case, multiply the two variables (radicands) and collect the product under the root:

$$\sqrt[n]{x} \times \sqrt[n]{z} = \sqrt[n]{xz} \text{ for } n > 1.$$

But, e.g.,

$$\sqrt[a]{x} \times \sqrt[b]{z} \neq \sqrt[ab]{xz} \text{ for } a \neq b,\ a, b > 1.$$

To see that this is so, observe that $\sqrt{25} \times \sqrt{9} = 5 \times 3 = 15 = \sqrt{225} = \sqrt{25 \times 9}$ because $15^2 = 225$. However, $\sqrt{25} \times \sqrt[3]{8} = 5 \times 2 = 10 \neq \sqrt[6]{200}$ because $10^6 \neq 200$.

Finding the quotient of two radicals is similar; one can simplify the quotient of two radicals only when their order is the same:

$$\frac{\sqrt[n]{x}}{\sqrt[n]{z}} = \sqrt[n]{\frac{x}{z}} \text{ for } n > 1.$$

But, e.g.,

$$\frac{\sqrt[a]{x}}{\sqrt[b]{z}} \neq \sqrt[a+b]{\frac{x}{z}} \text{ for } a \neq b,\ a, b > 1.$$

3.2.3.6 Other Functions

Of course, this small array of functions is not the entirety of those used in political science. One commonly used is the *absolute value*, which we denote by $|x|$. In a single dimension it just means "remove the sign on the value." More formally, it can be represented as $|x| = \sqrt{x^2}$ in one dimension, where we take only the positive root, or $|x| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$ in n dimensions, with $x = (x_1, x_2, \ldots, x_n)$. The absolute value is often used when one wants to keep the function positive (or negative with $-|x|$) over the entire range of x, or when one is interested in the distance between two points, which is $|a - b|$. Less commonly observed functions are rational functions (the ratio of two polynomials) and trigonometric functions (e.g., sine, cosine, tangent).[29]

Thus far all the functions we've defined have been the same over the entire domain. In other words, $f(x) = x^2$ doesn't change with the value of x. But we can also define functions *piecewise*, by which we mean simply "in pieces." These are useful, for example, when we expect an external intervention to alter the behavior of the relevant actors in a theory. There is nothing fancy to representing this sort of thing; we just write something like $f(x) = -(x-2)^2$ if $x \leq 2$ and $f(x) = \ln(x-2)$ if $x > 2$. This function states that below one's ideal point of 2, the function slopes downward at a faster rate than it slopes upward above one's ideal point.[30] One need only be careful to define the function across the entire domain, without missing some region. Piecewise functions are often expressed in the format

$$f(x) = \begin{cases} -(x-2)^2 & : x \leq 2, \\ \ln(x-2) & : x > 2. \end{cases}$$

3.2.3.7 Why Should I Care?

One encounters nonlinear functions both in formal theory and in statistics, as the examples sprinkled throughout this subsection demonstrate. We have already discussed the theoretical value of nonlinear functions: they provide us with a language to make very explicit statements about expected causal relations. And it turns out that exponential and log functions are useful for modeling and for transforming variables with highly skewed distributions, and that has both theoretical and statistical value, though this won't be very clear until we discuss distributions. Along these lines, the most used probability distribution, the normal distribution, is an exponential function composed with a quadratic.

[29] The trigonometric functions are rarely used in political science, but they can be important in situations in which they are used. Consequently, we include them in several of this book's discussions for reference but do not advise the first-time reader to worry about them.

[30] If this is not clear, draw the function. This is generally useful advice.

3.2.4 Why Should I Care?

Linear and nonlinear functions are nothing more than specific claims about the relationships among several variables, and thus can be very useful for making specific causal claims. Multivariate functions are especially useful as few political scientists suspect that much of politics can be usefully explained with bivariate hypotheses (i.e., conjectures that say only one concept is responsible for variation in another concept).

Having said that, let us briefly explain how one can move from verbal conjectures to writing down more specific functions. We will work with the probability of voting in a national election as an example. Suppose one suspects that the probability that a registered voter will cast a ballot will increase in response to (1) an individual's education level, (2) partisan identification, (3) income, (4) age, and (5) the closeness of the race.

Conceptualize education as a discrete count of the number of years of formal education, partisan identification as a distinction between those who identify with one of the two major parties and those who do not (we will assume this is a US election), income as continuous, age as a discrete count of years, and the closeness of the race as the gap between the Democratic and Republican candidates. We can represent the conjecture that the probability of voting is a function of each of these variables with the implicit function $p_v = f(ed, p, i, a, c)$, where p_v represents the probability of voting, ed represents education, p represents partisan identification, i represents income, a represents age, and c represents closeness. This equation is called the implicit functional form because it is not specific: we do not know whether the variables have positive, negative, linear, monotonic, or nonlinear effects on p_v. All the implicit function tells us is that they may have *some* effect. Hypotheses based on implicit functions are always more difficult to falsify than explicit ones that spell out the specific functional forms.

We might conjecture more strongly that the probability of casting a ballot is an affine function of each of these variables. That conjecture can be captured by the following function: $p_v = \alpha + \beta_1 ed + \beta_2 p + \beta_3 i + \beta_4 a + \beta_5 c$, where α is the intercept (i.e., the expected value of p_v when all of the explanatory variables have a value of 0) and the β_i parameters represent the strength of the impact that each explanatory variable has on voting probability.

Alternatively, we might conjecture that voting probability has a linear relation with some variables and a nonlinear relation with others. For example, one might argue that the impact of education is greatest at low levels (i.e., the difference between a fourth-grade and an eighth-grade education has a larger impact on voting probability than the difference between an eighth-grade and a twelfth-grade education, and the difference between completing a high school degree and completing a college degree has an even smaller impact). In addition, one might contend that greater levels of income have an even greater effect on voting probability. The following equation represents those conjectures: $p_v = \alpha + \beta_1 \ln(ed) + \beta_2 p + \beta_3 i^2 + \beta_4 a + \beta_5 c$.

The arguments presented in the preceding paragraphs are similar but distinct. By writing out an explicit functional form to represent the verbal arguments, one makes it very clear how the arguments are distinct (and how they are similar). Drawing graphs can often help one decide whether a given expected relationship comports with one's assumptions, intuition, or verbal argument. A functional representation of conjectures in equation form also makes very clear how one can be wrong—if statistical analysis shows that the parameters do not have the expected signs, for example—and that is another virtue of writing out the functional representation of a verbal argument.

Finally, one might conjecture an interactive relationship among some of the independent variables and probability of voting. There are many such possibilities, but, for example, one could believe that the higher one's level of education, the more the closeness of the race matters, as one will pay more attention to the media's reporting on the race. If this were true, we might expect a relationship like $p_v = \alpha + \beta_1 ed + \beta_2 p + \beta_3 i + \beta_4 a + \beta_5 c + \beta_6(ed \cdot c)$. Or one might suspect that the relationship between each explanatory variable and the explained variable grows with the value of the explanatory variable, and that the strength of each relationship is conditional on the values of the other explanatory variables.[31] One can represent such an argument as follows $p_v = \alpha \cdot ed^{\beta_1} \cdot p^{\beta_2} \cdot i^{\beta_3} \cdot a^{\beta_4} \cdot c^{\beta_5}$. We can take advantage of the relationship between exponents and logs to rewrite that as $\ln p_v = \ln(\alpha) + \beta_1 \ln(ed) + \beta_2(\ln p) + \beta_3 \ln(i) + \beta_4 \ln(a) + \beta_5 \ln(c)$.[32] Such a transformation is useful because while we cannot use common statistical routines to estimate the β parameters in the first representation, we can do so in the second representation. And while arguments that produce such a functional form are not usually observed in political science, they are in economics. It might be the case that few, if any, political processes are composed of concepts with such nonlinear, interactive relationships, but it might also be the case that few political scientists have explored those possibilities.

One can draw another illustration of the use of functions from game theory. In Chapter 1 we made reference to sets composing an actor's set of action. One can also create a set of strategic responses to all possible actions and all possible states of the world: a strategy set or strategy space. A strategy is a complete plan for playing a game (i.e., the choice an actor would make at each decision point the actor faces). So a strategy for player 2 might look like this: "if player 1 does x, then player 2 chooses a; if player 1 does y, then player 2 chooses b; etc." Strategies are functions (or correspondences): they map the relationship between the choices of the other players and the choices one makes at each opportunity. Strategies are sometimes represented as pairs of ordered sets rather than graphs or equations, but they are functions (or correspondences) nonetheless. Individual preferences in game theory also often take the form of

[31] We do not have a story to explain why such a conjecture is reasonable—it likely is not a reasonable conjecture. We offer it merely for illustrative purposes.

[32] If you found that too quick, observe that the first task is to take the log of all variables on both sides of the equation. The second step is to recall that $\ln(ed^b) = b \ln ed$.

functions, called utility functions. We provide an extended discussion of this example in the next section.

To reiterate the key point, one should develop a working familiarity with functional forms because they help one clarify the conjectures one is making. More specific causal claims are stronger because they are easier to falsify. Debates among scholars are also sharpened when there is greater clarity about the claims being advanced by the various factions. In short, good science becomes easier as clarity improves, and functions are a very basic and useful tool for adding clarity to one's conjectures.

3.3 PREFERENCE RELATIONS AND UTILITY FUNCTIONS

Game theory is a tool for understanding strategic interactions between political actors and developing theories about political behavior and the effect of institutions. The preferences of individual actors are foundational to game theory, as one cannot understand how one responds to incentives and others' actions without understanding what one actually wants. Typically, individual preferences are represented by functions, and the properties of these functions mirror the structure of one's preferences in the same manner that the form of the function described at the end of the last section matches one's theoretical expectations about the probability that one votes. In this section, we go into some detail as to why this is so, and how it all works, as an extended example of the usefulness of functions in political science. Before getting to functions, though, we need to return to relations.

3.3.1 Preference Relations

People frequently use a capital R to represent a relation, as follows: aRb, which is read "a is related to b." When applied to preferences, aRb is read "a is at least as good as b." If a and b were real numbers, this would translate to $a \geq b$; we return to this comparison below. There are also other analogues: aPb is "a is strictly preferred to b,"[33] or $a > b$ if both are numbers, and aIb is "I am indifferent between a and b,"[34] or $a = b$ if both are numbers. The study of these preference relations underlies decision theory, which, along with social choice theory, the study of group decision making, is often taught in parallel with or as a precursor to game theory. Many results from social choice theory are quite well known in political science. Black's median voter theorem, Arrow's (1950) impossibility theorem, and McKelvey's (1976) chaos theorem are notable examples you will likely be exposed to in other classes.

Our interest here is not in social choice heory, however, but rather in how to represent preferences with functions. To that end, we skip to a few important

[33] Formally, aPb if aRb but not bRa.
[34] Formally, aIb if aRb and also bRa.

propertiest that we often like preferences to have.[35] These are *completeness*, *transitivity*, *symmetry*, and *reflexivity*.

Completeness states that for any a and b, either aRb or bRa. In other words, all elements can be ordered pairwise, and there is no pair of elements for which one has no opinion. This is weaker than it may sound, as "no opinion" is distinct from indifference, which is allowed. What completeness disallows is the ability of someone to be unsure if she prefers a to b, b to a, or is indifferent between the two. For example, imagine a situation where a bureaucrat could (1) implement a new regulation (m), (2) implement the new regulation half-heartedly (h), or (3) ignore the new regulation (g). If the set is complete with respect to R, then one can have the preferences mRh, hRm, or both (mIh), but not neither. In other words, one can't say mPh and hPm simultaneously depending on mood, which is a formal way of denoting a lack of fixed opinion. Both the integers and the real numbers are complete relative to the normal ordering you are familiar with, given by the relations $>, \geq, =, \leq, <$. One never can be unsure whether $3 < 5$, for instance.

Transitivity states that if a is at least as good as b, and b is at least as good as c, then a is at least as good as c: if aRb and bRc, then aRc. The $>$, $\geq$, $<$, $\leq$, and $=$ relations are all transitive relations (e.g., if $a < b$ and $b < c$, then $a < c$) when applied to the integers or real numbers. To consider a political example, return to the set of implementation options for the bureaucrat: $\{m, h, g\}$. If she prefers ignoring the new regulation to implementing it half-heartedly, and also prefers implementing it half-heartedly to implementing it, then for her preferences to exhibit a transitive relation she would need to prefer ignoring it to implementing it.

Symmetry states that if aRb, then bRa for all a and b. In the realm of preference orderings, this implies complete indifference: everything is at least as good as everything else. The equality relation, $=$, is the only symmetric relation of $>, \geq, =, \leq$, and $<$ in the integers or real numbers: if $a = b$, then $b = a$.[36] Symmetric preference orderings are less common in the study of politics, though they do allow for a quite precise definition of the concept of "apathy," which otherwise might admit multiple interpretations. For instance, if our bureaucrat were indifferent between all three implementation options, then she would hold symmetric preferences. In this scenario, she would not only not care which option were chosen, but she would also be unlikely to put forth effort to affect the choice, assuming effort were at all costly.

Some people find **reflexivity** a bit of a brain bender. A relation on a set A is reflexive if for all $a \in A$, aRa is true. To illustrate, let's try the relations $>$, $\geq$, and $=$, and determine whether each is reflexive on the integers or real

[35] We state these in terms of preference relations, not more generally, as that is the only context in which we will have occasion to use them in this book. Note that these are normatively desirable properties, not properties observed to be true empirically. In fact, people violate these on a regular basis!

[36] While $a \geq b$ and $b \geq a$ might both be true (if a and b are equal), $a \geq b$ does not imply $b \geq a$ for all a and b.

numbers. To check the relation $>$, we replace the R in aRa with $>$ and see if it is true: $a \not> a$, so "greater than" is not a reflexive relation. However, $a \geq a$ and $a = a$ are both true, so "greater than or equal to" and "equal to" are reflexive relations. Now let's try a political science example. For our bureaucrat's preferences to exhibit reflexive order, each preference must be at least as good as itself: ignoring the new regulation must be at least as good as ignoring the new regulation, implementing the new regulation must be at least as good as implementing the new regulation, etc. We suspect you will agree that it would be odd indeed if someone's preferences were not reflexive.

3.3.2 Utility Functions

Complete and transitive individual preference is a fundamental assumption of rational choice theory and standard game theory, and is commonly assumed in the formal literature. It is true that people routinely violate this assumption in their everyday lives. However, the assumption buys us something very important—the ability to represent preferences with functions that take on real and integer values. To see why, let's return to the previous definitions. Integers and the real numbers are complete and transitive for all the usual ordering relations. Thus, if we want to represent our "at least as good as" relation with numbers, this relation had better have the same properties. With this assumption on individual (not group!) preference, we can translate the relation R on any set A to a function u on the same set. This u is called a *utility function* and assigns a value, typically a real number, to each element in A. So, for example, for a bureaucrat whose preferences are ordered $mRhRg$, we could assign $u(m) = 3, u(h) = 2$, and $u(g) = 1$.

This technique begins to pay dividends when the set of things one has a preference over is large, or infinite. For instance, while one could laboriously elaborate on preferences over dollar values of money ($100R99R98R\ldots$), it's far easier to define a utility function, $u(x) = x$, that represents those preferences. Varying the utility function alters what preferences are represented, in the same manner that varying the empirical model represents different theoretical ideas. A linear utility like $u(x) = ax$ for budgetary outlays, for example, would mean that each additional dollar is just as valued as the previous one. A quadratic utility like $u(x) = ax^2$, in contrast, would mean that each additional dollar is valued more than the one before it, an unlikely assumption in many cases (though see Niskanen, 1975, p. 619). In fact, for money, researchers typically assume that $u(x) = \ln(x)$, so that there are decreasing returns to increasing a bureaucratic budget.

This makes sense in the context of a single agency's preferences, but what about a Congress trying to distribute money over multiple agencies? Each congressperson might have some ideal budget number for each agency, with increases *and* decreases from that number being viewed negatively. In that case, we can use what is known as a quadratic loss function, $u(x) = -(x - z)^2$. If you graph this function, you will see it is a parabola that peaks at $x = z$, which is the point of highest utility, also known as an ideal point. This form of utility function is very common when modeling voting behavior (e.g., McCubbins, Noll, and Weingast, 1987).

3.3.3 Best Response Correspondences

Let's return to the example of the bureaucrat, but now assume there are two decision makers. One, Christine, has preferences $mPhPg$. She prefers to do it right, but also wants it done. The other, Bob, is lazy and has preferences $gPhIm$. He'd rather do nothing, but if it has to be done, he doesn't care which way it happens. Let's also assume that, for some unknown reason, the decision is made by asking Christine and Bob to write their choices on a piece of paper. If both agree, then that option wins. If only one writes m, then h happens. In this (odd) scenario, Christine will always write m. This is a dominant strategy for her, because it can secure her second-best option and possibly achieve her first-best option. Bob, on the other hand, is in a pickle. He can't get his best option given Christine's optimal action, and he is indifferent between h and m. Thus anything he does has the same outcome to him. His best response to Christine is any of the three options.

We can represent Christine's best response as a function. Let $S = \{m, h, g\}$, which is known as a player's strategy space. Then we can write the function $B_C(\cdot) = m$ for Christine, which means that her best option is to choose m regardless of what Bob does. To elaborate, $B_C(m) = B_C(h) = B_C(g) = m$. B_C here is called Christine's **best response function**. It takes as input Bob's strategies and returns the optimal action for Christine to take. It is a function because Christine has only one best response to each of Bob's actions.

Now consider Bob's best response to Christine's play of m. We can't represent this best response as a function, as it would have to return three values—m, h, or g—when presented with Christine's m. Instead, we can write Bob's **best response correspondence**. Formally, Bob's decisions are governed by the correspondence $B_B(m) = \{m, h, g\}$. In words, this means that Bob responds optimally to Christine's choice of m by choosing any of his options. We write such correspondences as $B_B(s_C) : S_C \rightarrow\rightarrow S_B$ where we have added subscripts for each player's name, and S_i and s_i are the strategy space and strategy choice for player i. Though we will not deal with correspondences much in this book, they will come up in your game theory classes.

3.3.4 Why Should I Care?

Whether or not they are your cup of tea, formal theories of political science are prevalent in the field and often referenced in empirical work to justify hypotheses. Being able to read them and understand their underlying assumptions are important skills. Further, formalizing theories can help sharpen your thinking. Finally, in the same manner that different utility functions represent different preferences, one can choose different underlying properties on preferences if one does not like, for instance, rational choice assumptions.

3.4 EXERCISES

1. For each pair of ordered sets, state whether it represents a function or a correspondence:

   a) {5, -2, 7}, {0, 9, -8}
   b) {3, 1, 2, 6, -10}, {5, 7, 1, 4, 9}
   c) {3, 7, -4, 12, 7}, {8, -12, 15, -2, 17}

2. Simplify $h(x) = g(f(x))$, where $f(x) = x^2 + 2$ and $g(x) = \sqrt{x - 4}$.
3. Simplify $h(x) = f(g(x))$ with the same f and g. Is it the same as your previous answer?
4. Find the inverse function of $f(x) = 5x - 2$.
5. Simplify $x^{-2} \times x^3$.
6. Simplify $(b \cdot b \cdot b) \times c^{-3}$.
7. Simplify $((qr)^\gamma)^\delta$.
8. Simplify $\sqrt{x} \times \sqrt[5]{x}$.
9. Simplify into one term $ln(3x) - 2ln(x + 2)$.
10. Visit the "Graphing Linear Functions" page at the Analyze Math website `http://www.analyzemath.com/Graphing/GraphingLinearFunction.html`. Read the examples and solve the two "matched problems."
11. Visit the Analyze Math website's "Slope Intercept Form of a Line" page at `http://www.analyzemath.com/Slope_Intercept_Line/Slope_Intercept_Line.html`. Print a copy of the page and then click on the Click to Start Here button to start the tutorial applet. Do numbers 2 through 8. What does this tutorial show?
12. Visit the "Quadratic Function(General Form)" page at Analyze Math: `http://www.analyzemath.com/quadraticg/quadraticg.htm`. Click on the Click Here to Start button and adjust parameters a, b, and c. What happens to the graph as you increase or decrease a? Note the changes when you increase b and c as well. Is there a value to which you can set one of the parameters to make the quadratic function a linear function?

13. Visit the "Graphs of Basic Functions" page at the Analyze Math site (`http://www.analyzemath.com/Graph-Basic-Functions/Graph-Basic-Functions.html`). Click on the Click Here to Start button and plot the graph of each function. After plotting each once, click the Y-Zoom Out button several times and plot each of the graphs again.

14. Visit "Polynomial Functional Graphs" at `http://id.mind.net/~zona/mmts/functionInstitute/polynomialFunctions/graphs/polynomialFunctionGraphs.html`. Plot polynomial functions of different orders, then adjust the parameters and observe how the graph changes in response to different values (use the Simple Data Grapher from the Math link on the main page). Write down a verbal conjecture about politics that you think might be captured by a specific polynomial function. Be sure to explain your thinking. Write down the function and print a copy of its graph.

15. Rewrite the following by taking the log of both sides. Is he result a linear (affine) function?
$y = \alpha + x_1^{\beta_1} + \beta_2 x_2 + \beta_3 x_3$.

16. Rewrite the following by taking the log of both sides. Is the result a linear (affine) function?
$y = \alpha \times x_1^{\beta_1} \times x_2^{\beta_2} \times x_3^{\beta_3}$.

17. Rewrite the following by taking the log of both sides. Is the result a linear (affine) function?
$y = \alpha \times x_1^{\beta_1} \times \frac{x_2^{\beta_2}}{x_3^{\beta_3}}$.

18. Is this problem done correctly? Yes or no.
Take the log of both sides of the following equation:
$y = x_1^{\beta} - x_2^{n} + x_3^{2}$.

19. Visit "The Universe Within" page on the website of Florida State University's magnet lab: `http://micro.magnet.fsu.edu/primer/java/scienceopticsu/powersof10/`. It is a visual display of the concept of scale—viewing the same object from different scales of measurement—as it begins with a view from 10^{+23} meters away and moves to 10^{-16} meters away. Besides being a cool visual, this page offers a graphic illustration of exponentiation.[37] Note especially what happens when the exponent shifts from positive to negative values.[38] If that does not make sense to you, review the discussion of exponents, specifically the arithmetic rules.

[37] We have purposely used political rather than physical examples throughout, but could not resist this one.

[38] Once it has run you may want to click on the Increase button to go back through it as it moves fairly quickly.

20. The graduate studies committee has asked the graduate students to provide the faculty with a list of three nominees to represent the students on the committee. After much discussion, three nominees are put forward, and you are asked to rank them, with the rank representing preference (i.e., 1 is most preferred, 2 is second best, and 3 is the third-best choice). The nominees are Beta, a seventh-year student who recently defended his prospectus; Gamma, a second-year student who is very bright but tends to dominate seminar discussion; and Alpha, a fourth-year year student who is preparing for her exams and is widely viewed as level-headed and realistic. Provide your pairwise preference rankings of each candidate. Check to see whether your rankings are transitive. If you have been assigned this problem for class, bring your ordering to class so that the class can determine whether it is transitive at the aggregate level under pairwise majority rule.

21. Recall the first question of Section 8.2 in Chapter 1. There we asked you to pick ideal spending points for three parties, as well as a status quo and a bill, and conjecture about whether or not it would pass. Now we want you to go further. Write a utility function for each party that is largest at that party's ideal point. How does that function decrease with distance from the ideal point? Try to draw a curve around each ideal point that gives the same utility to the corresponding party for every point on the curve. These are called *indifferences curves*, as the party is indifferent between all points on the curve. Draw these for all parties and see whether you can answer your earlier conjecture.

22. Propose and justify a quadratic utility function as representing the preferences of some political actor over something.