International Journal of Science Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsed20

The nature and development of hypothetico-predictive argumentation with implications for science teaching

Anton Lawson [a]

[a] School of Life Sciences, Arizona State University, Tempe, AZ 85287-1501, USA; e-mail: anton.lawson@asu.edu

Published online: 03 Jun 2010.

To cite this article: Anton Lawson (2003): The nature and development of hypothetico-predictive argumentation with implications for science teaching, International Journal of Science Education, 25:11, 1387-1408

To link to this article: http://dx.doi.org/10.1080/0950069032000052117

INT. J. SCI. EDUC., NOVEMBER 2003, VOL. 25, NO. 11, 1387–1408

RESEARCH REPORT

The nature and development of hypothetico-predictive argumentation with implications for science teaching

Anton E. Lawson, School of Life Sciences, Arizona State University, Tempe, AZ 85287–1501, USA; e-mail: anton.lawson@asu.edu

This paper explicates a pattern of scientific argumentation in which scientists respond to causal questions with the generation and test of alternative hypotheses through cycles of hypothetico-predictive argumentation. Hypothetico-predictive arguments are employed to test causal claims that exist on at least two levels (designated stage 4 in which the causal claims are perceptible, and stage 5 in which the causal claims are imperceptible). Origins of the ability to construct and comprehend hypothetico-predictive arguments at the highest level can be traced to pre-verbal reasoning of the sensory-motor child and the gradual internalization of verbally mediated arguments involving nominal, categorical, causal and, finally, theoretical propositions. Presumably, the ability to construct and comprehend hypothetico-predictive arguments (an aspect of procedural knowledge) is necessary for the construction of conceptual knowledge (an aspect of declarative knowledge) because such arguments are used during concept construction and conceptual change. Science instruction that focuses on the generation and debate of hypothetico-predictive arguments should improve students' conceptual understanding and their argumentative/reasoning skills.

Introduction

Newton et al. (1999) make a compelling case for increasing student argumentation in the pedagogy of school science. In brief, their position is that argumentation functions in science to construct plausible links between imaginative conjectures of scientists and evidence. Thus, a central activity of the scientist is to construct and use arguments about which of the imaginative conjectures for a puzzling phenomenon are the most convincing in light of that evidence and, of course, to obtain additional evidence when the available evidence is insufficient or lacking. Because argumentation plays such a central role in science, Newton et al. (1999) propose that argumentation should also play a central role in science pedagogy. The presumption is that engaging students in argumentation is the best way for them to not only construct conceptual knowledge, but also to become skilled in the use of general forms of argumentation that are of considerable use in democratic societies. The purpose of the present paper is to expand on this theme by explicating the nature of hypothetico-predictive argumentation, its development, and its potential use in the science classroom. A subsequent paper will discuss our efforts at putting this theme into action in a college biology course and evaluating the outcomes in terms improved student reasoning/argumentative skills.

International Journal of Science Education ISSN 0950–0963 print/ISSN 1464–5289 online
http://www.tandf.co.uk/journals
DOI: 10.1080/0950069032000052117

Toulmin's framework of scientific argumentation

How do scientists construct and use arguments? Of course, several forms of argumentation (e.g. sign, analogy, methods of difference, agreement and concomitant variation) have been identified (for example, Freeley 1976, Hiatt 1975, Lawson and Kral 1985, Olson 1969, Salmon 1995). Nevertheless, like Newton et al. (1999), let us start within the framework proposed by Toulmin (1958). In Toulmin's framework, at the heart of scientific argumentation lie *claims* that the scientist wishes to establish as correct and the *evidence* (data) that one appeals to as a foundation for the claims. In addition to claims and evidence, general standards or cannons of argument, called *warrants*, exist that link claims with evidence. Finally, statements exist that function essentially as assurances that the warrants are in fact valid. Toulmin refers to these statements as *backings*. Toulmin (1958) describes scientific argumentation primarily as a process of using evidence and warrants and backings to convince others of the veracity of specific claims. For example, figure 1 (after Toulmin et al. 1984: 338) depicts these elements in an argument designed to convince others that goiter is caused by iodine deficiency.

The general pattern of hypothetico-predictive argumentation

Although convincing others of the validity of causal claims is certainly an important component of science, the present view of scientific argumentation differs

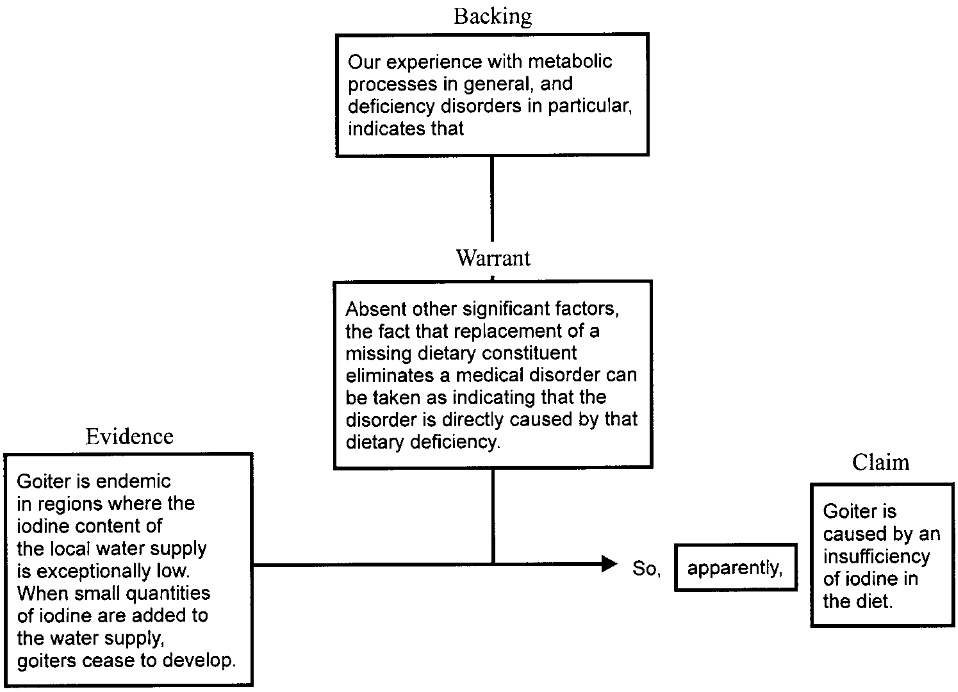

Figure 1. The evidence, warrant and backing embedded in an argument designed to support the claim that goiter is caused by an iodine deficiency (after Toulmin et al. 1984: 338).

somewhat from that proposed by Toulmin in that it emphasizes argumentation in the process of testing alternative hypotheses. In other words, the initial goal of argumentation in the present view is to discover which of two or more proposed alternative explanations (claims) for a puzzling observation is correct *and* which of the alternatives are incorrect. In this sense, a claim can be thought of not so much as something that a particular scientist (or science student) believes is correct at the outset. Rather, a claim can be thought of as a tentative explanation that *may* be correct, thus is in need of test through the generation of specific predictions and the gathering of evidence. Of course once scientists, or science students, reach a sound conclusion (i.e. have 'discovered' the 'correct' explanation), they must once again engage in argumentation to convince others of the veracity of their conclusion. In theory, this should amount to constructing arguments *for* one of the alternatives and *against* the others. Nevertheless, in both Toulmin's view and in the present view, scientific arguments contain claims and evidence, and may also contain warrants and backings.

Testing alternative hypotheses: why do people use spices?

To introduce the present pattern of scientific argumentation, consider a recent article in the *American Scientist* by Sherman and Flaxman (2001). In that article, the authors noted that humans have used several kinds of spices throughout history and wondered why. In their words:

> Black pepper, for example, is the world's most widely used spice even though *Piper nigrum* grows naturally only in the New World tropics. What accounts for the enduring value of spices? The obvious answer is that they enhance flavor, color and palatability of food. This proximate, or immediate-cause, explanation is true but it does not address ultimate, or long-term, questions of why people find foods more appealing when they contain pungent plant products, why some secondary compounds or tastier than others and why preferences for these chemicals differ among cultures. (2001: 142–143)

After pointing out that the ultimate reason why plants possess secondary compounds is disease protection, Sherman and Flaxman advanced an ultimate hypothesis for human spice use. Again, in their words:

> Our foods are also attacked by bacteria and fungi, often the same ones that afflict the spice plants. If spices were to kill microorganisms or inhibit their growth or production of toxins, then spice use could protect us from food-borne illnesses and food poisoning. (2001: 143)

Sherman and Flaxman continued by discussing how they tested this hypothesis:

> In a series of recent studies, one of us [Sherman], along with Jennifer Billing and Geoffrey Hash, set out to test this 'anti-microbial hypothesis.' We located 107 'traditional' cookbooks from 36 countries, representing every continent and 16 of the world's 19 major linguistic groups. To test the anti-microbial hypothesis, we developed five critical predictions and examined them by combining information from the microbiology literature with analyses of traditional recipes. (2001: 143)

The five critical predictions were presented as follows:

> Prediction 1: Spices used in cooking should exhibit anti-microbial activity. (2001: 143)
> Prediction 2: Use of spices should be greatest in hot climates, where un-refrigerated foods spoil quickly. (2001: 144)

> Prediction 3: The spices used in each country should be particularly effective against the local bacteria. (2001: 145)
> Prediction 4: Within a country, meat recipes should be spicier than vegetable recipes. (2001: 145)
> Prediction 5: Within a country, recipes from lower latitudes and altitudes should be spicier because of the presumably greater microbial diversity and growth rates in these regions. (2001: 145)

Not only did Sherman and Flaxman present these predictions, they also stated a warrant needed to tie the predictions to their hypothesis:

> The anti-microbial hypothesis assumes that the amounts of spices called for in a recipe are sufficient to produce the desirable effects and that cooking does not destroy the active chemicals. Although the efficacy of spices in prepared meals has not been evaluated directly, both assumptions seem reasonable. The minimum concentrations of purified phytochemicals necessary to inhibit growth of food-borne bacteria *in vitro* are well within the range of spice concentrations used in cooking. Phytochemicals are generally thermostable, and spices containing those that are not, such as cilantro and parsley, are typically added after cooking, so their anti-microbial effects are not lost. (2001: 144)

The evidence reported by Sherman and Flaxman need not be detailed. It suffices to say that, once gathered, it matched the five predictions extremely well. For example, with respect to Prediction 2, Sherman and Flaxman reported:

> In five of the six hottest countries (India, Indonesia, Malaysia, Nigeria and Thailand) every meat-based recipe called for at least one spice, whereas in the two coldest countries (Finland and Norway) more than a third of the meat-based recipes did not call for any spices at all. (2001: 144)

Perhaps because the reported evidence so strongly supported the anti-microbial hypothesis, Sherman and Flaxman did not bother to say so, other than stating that the frequency of spice use in meat-based recipes '. . . is consistent with their presumptive role in inhibiting food-borne bacteria' (2001: 146) Nevertheless, one hypothetico-predictive argument embedded in the article can be summarized in the following way:

> *If* . . . spices protect us by killing microbes in food [anti-microbial hypothesis],
> *and* . . . meat recipes in traditional cookbooks of several hot and cold climate countries are analyzed for spice use [planned test],
> *then* . . . spice use should be greater in hot climate countries where un-refrigerated foods (particularly meats) spoil quickly [prediction].
> *And* . . . in five of the six hottest countries every meat-based recipe called for at least one spice, whereas in the two coldest countries more than a third of the meat-based recipes did not call for any spices at all [observed result].
> *Therefore* . . . the anti-microbial hypothesis is supported [conclusion].

The major elements of hypothetico-predictive argumentation

Using this argument as an example, figure 2 depicts the major elements of hypothetico-predictive argumentation as well as the key words that link the elements together. The perceptible elements appear as solid boxes while the imagined elements appear as 'clouds'. As shown, the process begins with a puzzling observation that provokes a causal question and then the generation of one or more tentative explanations (alternative hypotheses – alternatives that often parallel explanations generated by previous generations of scientists; cf. Wandersee et al. 1994). As discussed later, hypothesis generation is a creative, open-ended activity. Thus, the list of generated hypotheses may in fact not contain a hypothesis that may

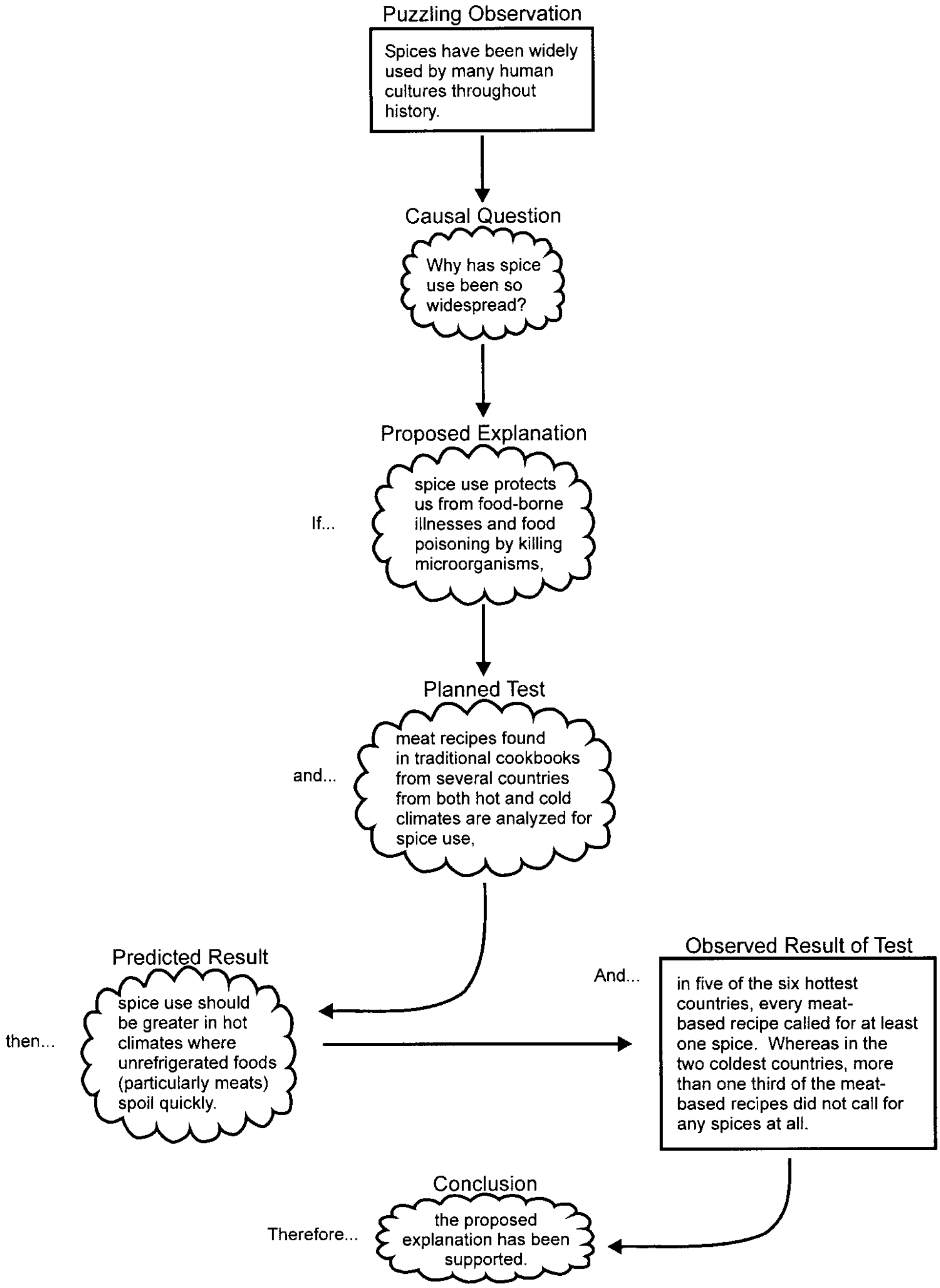

Figure 2. A hypothetico-predictive argument used to test the anti-microbial hypothesis.

eventually be found 'correct'. Regardless of this limitation, once generated, they must be tested. To test a hypothesis, one begins by assuming that the hypothesis is correct. Next, one must imagine a test that together with the hypothesis *should* produce one or more specific observable results (the prediction). The words *If/and/*

then link the hypothesis and imagined/planned test to the prediction. Logicians refer to the *If/and/then* reasoning used to generate a prediction as *deduction* (e.g. if A > B and B > C, then it follows via deduction that A should be > C) (for example, Conway and Munson 1990, Olson 1969, Salmon 1995). However, because generating predictions may involve a creative element (i.e. more than an automatic and 'logical' act of deduction), the overall pattern of argumentation depicted in figure 2 is labeled *hypothetico-predictive*, rather than hypothetico-deductive (cf. Lewis 1988, Medawar 1969, Popper 1959, 1965). Nevertheless, once a test is planned, complete with one or more predictions, the test is conducted. The test then yields observed results. The observed results constitute evidence. Finally, the evidence is compared with the prediction. The match or mismatch of evidence and prediction is then used to draw a conclusion regarding the degree of support or non-support for the hypothesis.

Testing alternative hypotheses

The main objective of the hypothetico-predictive argument depicted in figure 2 is to test the antimicrobial hypothesis. Of course testing and finding support for one hypothesis does not constitute non-support for alternative hypotheses. Indeed, the present view calls not only for the construction hypothetico-predictive arguments in favor of one hypothesis, but also for hypothetico-predictive arguments relevant to the alternatives. With this in mind, let us return to the Sherman and Flaxman article to discover what they did next. Again in their words:

> One can imagine other reasons why spices may have been so common in the human diet. Perhaps the most prominent of these suggests that spices disguise the smell and taste of spoiled foods . . . Another possible idea suggests that spices might serve as mendicants . . . It has also been suggested that spicy foods might be preferred in hot climates because they increase perspiration, thus help cool the body . . . Finally, it could be that people use whatever aromatic plants grow locally just because they taste good. (2001: 145–146)

Thus, Sherman and Flaxman did not simply test the anti-microbial hypothesis. They also discussed several alternatives. For example, this is what they wrote about the hypothesis that people use spices because they taste good:

> Under this proximate-level hypothesis, spice chemicals should be highly palatable, and spice-use patterns should correspond to availability. Neither prediction is fully supported. There is no relation between the number of countries in which each spice plant grows and either the number of countries in which it is used or their annual temperatures. Second, pungent spices like garlic, ginger, anise and chilies are initially distasteful to most people. (2001: 146)

Accordingly, we should be able to cast these statements in the form of a hypothetico-predictive argument. As shown in the following, this is easily accomplished:

> *If* . . . spices are used because they taste good [taste-good hypothesis],
> *and* . . . the palatability and availability of spices are assessed [planned test],
> *then* . . . spice chemicals should be highly palatable and spice-use patterns should correspond to availability [prediction].
> *But* . . . spices such as garlic, ginger, anise and chilies are initially distasteful to most people and there is no relationship between where spices are used and where they grow [observed result].
> *Therefore* . . . the taste-good hypothesis is not supported [conclusion].

To summarize, a primary role of scientific research is to explain puzzling observations by generating hypothetico-predictive arguments for and against as many alternative explanations that the researcher, or other researchers, are able to generate. Of course some explanations are more plausible than others in terms of what is known in related fields of inquiry. Thus, it is reasonable to test the most plausible explanations first. After all, in spite of the claim that science progresses best via falsification (for example, Popper 1959), one does not receive a Nobel Prize for being wrong or for demonstrating that others are wrong. Nevertheless, the main point is that the research process is not complete until convincing hypothetico-predictive arguments have been advanced concerning each of the reasonable alternatives. Of course, the claim here is not that all scientists are consciously aware of the identified pattern of hypothetico-predictive argumentation. Perhaps some are. Perhaps others are not. Certainly it would seem that conscious awareness of the pattern would at least help the scientist present his/her research to others, if not improve the quality of the research in the first place (for example, Chamberlain 1965, Platt 1964, Lawson 2000a, 2000b).

The creative aspects of hypothesis and prediction generation

As already mentioned, both hypothesis and prediction generation involve creative elements. Within the psychological literature, hypothesis generation is understood as a process involving analogies, analogical transfer, or analogical reasoning (i.e. borrowing ideas that have been found to 'work' in one or more past related contexts and using them as possible solutions/hypotheses in the present context) (cf. Biela 1993, Boden 1994, Bruner 1962, Dreistadt 1968, Finke et al. 1992, Gentner 1989, Hestenes 1992, Hoffman 1980, Hofstadter 1981, Holland et al. 1986, Johnson 1987, Koestler 1964, Sternberg and Davidson 1995, Wong 1993). Within the more traditional literature of speech, communication and English, analogy use is commonly referred to as a form of argumentation known as argument by analogy (for example, Warnick and Inch 1989). The use of analogical reasoning, or argument by analogy, is a creative act dependent in part on background declarative knowledge. One of the most well-known examples of the creative element involved in generating a hypothesis is that of Friedrich von Kekule. Kekule dosed off one afternoon in 1865 and dreamt of a snake twisting around and seizing hold of it own tail. According to Kekule, this dream provoked him to generate the hypothesis that certain organic molecules, such as benzene, are not open structures but closed rings, like the snake biting its tail (as quoted in Koestler 1964).

Prediction generation also involves an element of creativity. A dramatic example of this can be seen in the research of Otto Loewi. For several years, Loewi had suspected that neural impulses were chemically transmitted from neurons to muscles. However, he was unable to think of a way to test this chemical-transmission hypothesis. Finally, one night in 1920, he literally dreamed up an experiment, complete with a prediction, that would do the trick. When he awoke the next morning, he immediately went to his laboratory and conducted the test. And, to his delight, the observed results turned out just as predicted, providing support for his chemical-transmission hypothesis and eventually winning him a Nobel Prize (cf. Koestler 1964, Lawson 2000a).

Of course these examples do not imply that we need to fall asleep and dream to be creative. But they do imply that a creative, non-conscious, element is involved in both hypothesis and prediction generation. Hence, successful hypothetico-predictive reasoning requires reflective thought and may take considerable time or may not happen at all. In the classroom, this implies that teachers should be willing to give students plenty of time to think and should be open to unusual and unexpected ideas.

Developmental stages and hypothetico-predictive argumentation

Research with adolescents and adults indicates that many experience considerable difficulty in constructing hypothetico-predictive arguments in theoretical contexts such as that investigated by Sherman and Flaxman (for example, Lawson et al. 2000a,b). Nevertheless, research also suggests that a rudimentary form of *If/and/then* reasoning is present virtually at birth. Accordingly, developmental changes may not involve changes in this reasoning pattern with age. Instead, they may involve changes in the contexts to which the reasoning pattern can be applied. Let us see how this might work in terms of five developmental stages that correspond in a general way to Piaget's four stages (for example, Inhelder and Piaget 1958, Piaget and Inhelder 1969).

Stage 1: the sensory-motor stage (birth–18 months)

Of course, children during the first 18 months of life do not generate *If/and/then* verbal arguments. Nevertheless, their overt behavior suggests that their pre-verbal reasoning follows this pattern. Consider, for example, Piaget's famous object permanence task in which an experimenter, in full view of the infant, hides a ball under one of two covers. Diamond (1990) has shown that infants as young as 5 months old will reach under the cover for the hidden ball, indicating that they retain a mental representation of the ball even though it is out of sight. Furthermore, such behavior suggests that the infant is reasoning in the following way:

> *If* . . . the ball is still where he/she put it, even though I can no longer see it [empirical representation],
> *And* . . . I reach under the cover where it was hidden [planned test],
> *Then* . . . I should find the ball [prediction].

In agreement with Meltzoff (1990), the infant's representation is labeled empirical because it is of an event that has been empirically experienced. That is, the infant actually saw the ball hidden under the cover.

Stage 2: the preoperational or 'nominal' stage (18 months–7 years)

Although infants younger than 18 months old solve a simple object permanence task, not until 18 months do they solve one in which they must represent what is not experienced. To tap this higher-order skill, Piaget (1954) invented a hiding task called serial invisible displacement. In this task, the adult hides a ball in his/her hand in view of the infant, and then moves the hand under a series of three occluders dropping the ball under one of them. The infant is not given any indication that the ball has been dropped off. Instead, the infant sees the hand emerge at the other side of the occluders. Consequently, the infant looks in the empty hand and must reason that, because the ball is not there, but must be somewhere. Thus it must be under one of the occluders.

The results of this experiment are that children younger than 18 months old look in the hand and look no further. They are stumped. However, after 18 months, children find the ball. Again, the reasoning seems to follow the *If/and/then* form:

> *If* . . . the ball is hidden behind one of the occluders [imagined representation],
> *and* . . . I lift each in turn [planned test],
> *then* . . . I should eventually find the ball [prediction].

Thus, what separates the first and second stage does not appear to be the *If/and/then* pattern. Rather, what seems to separate the stages is the context in which the pattern can be applied. In this second, more difficult task, the child initiates reasoning with an imaginary, as opposed to an empirical, representation. In Meltzoff's words:

> By 18 months of age there has been the growth of a kind of second-order representational system and a capacity for hypothetical representations. This enables the child to wonder 'what if,' to contemplate 'as if,' and to deduce 'what must have been' in advance of, and often without, the perceptual evidence. (1990: 22)

A major achievement during this stage is the acquisition of language and its use in the naming of objects, events and situations constructed during this and the prior stage.

Stage 3: the concrete operational or 'categorical' stage (7 years–early adolescence)

The acquisition of language to name objects, events and situations during stage 2 now allows for application of the *If/and/then* pattern to a new level, the level of seriating and classifying (i.e. creating higher-order classes/categories of imperceptible objects, events and situations). The perceptible and named objects of stage 2, such as tables and chairs, become the imperceptible categories, such as furniture, of stage 3.

Lawson (1993b) administered a series of tasks, including the Mellinark task (figure 3), to children ranging in age from 6 to 8 years. Carefully sequenced instruction was used to teach the children how to use *If/and/then* reasoning to discover the relevant attributes of the creatures:

> *If* . . . tiny spots are the key feature that makes a creature a Mellinark [categorical proposition],
> *and* . . . I look at all of non-Mellinarks in row 2 [planned test],
> *then* . . . none should have tiny spots [prediction].

Interestingly, none of the 30 6 year olds were able to generate and/or comprehend this sort of argument when verbally presented, whereas 15 of the 30 7 year olds were, as were virtually all of the 8 year olds (29 of 30). Levine and Prueitt (1989) and Dempster (1992) reviewed research indicating that younger children's failure may be related to relatively late maturation of the prefrontal cortex.

The present position is that stage 3, which beings at age 7, involves use of the *If/and/then* pattern to order and categorize the objects, events, and situations in the

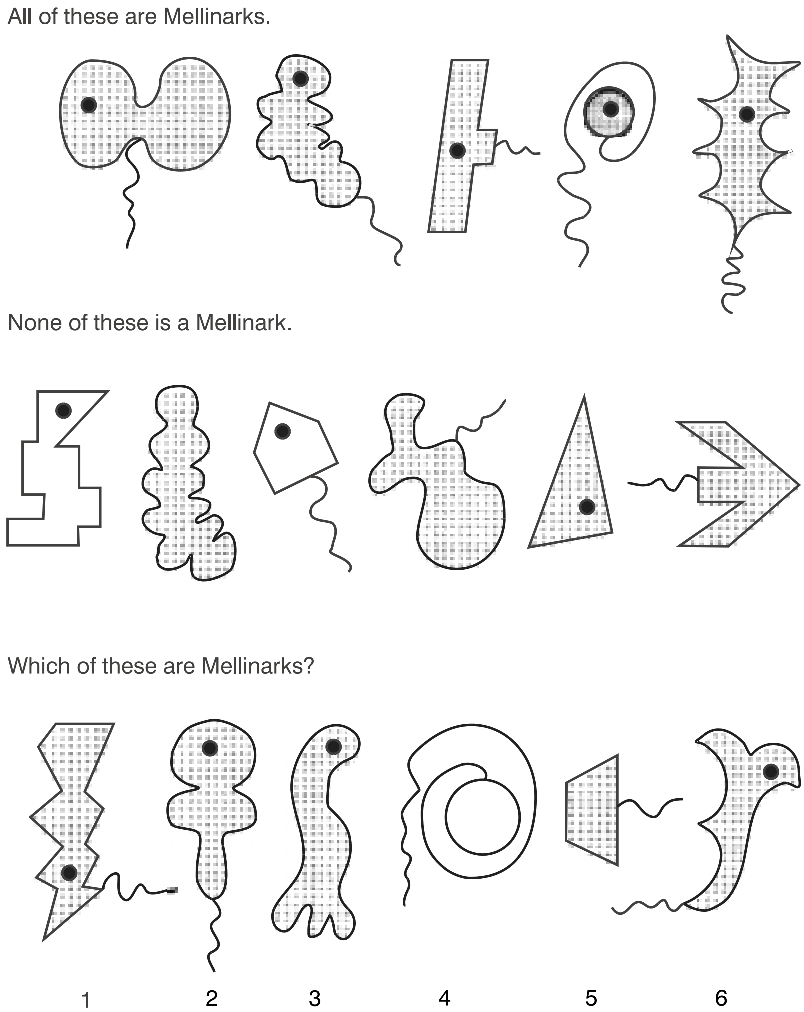

Figure 3. Which creatures in row three are Mellinarks? The Mellinark task.

child's environment all mediated by language. However, there is a distinct limitation. Reasoning (like that in stage 1) is initiated with what the child perceives (e.g. the child actually sees the tiny spots on the creatures in the Mellinark task). In this sense, the representations used to initiate reasoning are empirical.

Stage 4: the formal operational or 'causal' stage (middle–late adolescence)

At roughly age 11–12 years, some adolescents become increasingly able to use language to apply *If/and/then* arguments to causal, rather than categorical, propositions. Consider, for example, the causal question: What causes differences in the rates at which pendulums swing? To answer this question, one must generate and test alternative causal hypotheses (cf. Inhelder and Piaget 1958):

> *If* . . . changes in swing rates are caused by the amount of weight hanging on the end [causal proposition],
> *and* . . . the weights are varied while holding other possible causes constant [planned test],
> *then* . . . rate of pendulum swing should vary [prediction].

Again, the *If/and/then* pattern is the same as in the prior three stages. Thus, again, the difference between stage 4 reasoning and prior reasoning is not the pattern, but to what the pattern is applied. Stage 3 reasoning is about testing categorical propositions (sometimes referred to as generalizing hypotheses). Stage 4 reasoning is about testing causal propositions (sometimes referred to as causal hypotheses). The already presented stage 4 planned test involves an experiment in which the possible cause is manipulated. In other words, the proposed cause is the amount of weight, and the experiment's independent variable is also the amount of weight. Importantly, this variable can be easily manipulated because weight differences can be directly sensed.

Stage 5: the post-formal or 'theoretical' stage (late adolescence and early adulthood)

Consider once again the Sherman and Flaxman investigation with its central causal question: Why has spice use been so extensive throughout human history? Recall that their anti-microbial hypothesis was tested using the following hypothetico-predictive reasoning:

> *If* . . . spices protect us because their chemicals kill microbes in food [theoretical proposition],
> *and* . . . meat recipes in traditional cookbooks of several hot and cold climate countries are analyzed for spice use [planned test],
> *then* . . . spice use should be greater in hot climate countries where un-refrigerated foods (particularly meats) spoil quickly [prediction].

Although identical to the prior stage 4 reasoning in form and intent, this argument differs in at least two important ways. First, here the proposed cause (i.e. unseen chemicals in spices killing unseen microbes) is non-perceptible. Furthermore, unlike the non-perceptible categories of stage 3 (e.g. furniture), here the non-perceptible entities have no perceptible exemplars (i.e. one cannot point to a carbon and a hydrogen atom as exemplars of elements in the way that one can point a chair and a couch as exemplars of furniture). Before the proposed cause (i.e. weight difference) was perceptible. And second, unlike stage 4 reasoning where a proposed cause and the independent variable of an experiment designed to test it were one and the same, this is no longer the case. In the earlier test, the independent variable is meat recipes in hot versus cold climates, while the proposed cause is the non-perceptible and imagined process of microbes being killed by the chemicals in spices. Because the proposed cause and the independent variable are not the same, a warrant (a theoretical rationale) must be generated to link the two so that a reasonable test can be conducted. For these reasons, stage 5 reasoning, labeled post-formal or theoretical, is more difficult and presumably not achieved until late adolescence, if at all (for example, Lawson et al. 2000a,b).

The internalization of patterns of external argumentation

How do individuals progress from one stage to the next? Piaget advanced the hypothesis that the development of reasoning occurred as a consequence of 'the shock of our thoughts coming into contact with others, which produces doubt and the desire to prove' (1928: 204) Piaget went on to state:

> The social need to share the thought of others and to communicate our own with success is at the root of our need for verification. Proof is the outcome of argument . . . Argument is therefore, the backbone of verification. Logical reasoning is an argument which we have with ourselves, and which produces internally the features of a real argument. (1928: 204)

This hypothesis seems consistent with that of Vygotsky (1962) who viewed speech as social in origin, and only with time did it come to have self-directive properties that eventually resulted in internalized thought. Luria (1961) also advanced a similar view. According to Luria, the progressive differentiation of language to regulate behavior occurs in four steps. First, the child learns the meaning of words. Second, language serves to activate, but not limit, behavior. Third, language controls behavior through activation or inhibition via communication from an external source. And fourth, the internalization of language serves a self-regulating function through instructions to oneself.

Later Piaget (1976) proposed a three-stage theory of operative or procedural knowledge development. The first stage (sensory-motor) is one in which language plays little or no role as it has yet to be acquired. The child learns primarily through sensory-motor activity and knowledge is that of action. The second stage is characterized by the acquisition of language. The child responds to spoken language and acquires knowledge transmitted from adults who speak the same language. To learn, the child raises questions and adults respond verbally to those questions. Of course, this is not to say that all adult responses are understood. Nonetheless, a new and powerful mode of learning becomes available. The essential limitation of this stage is that the use of language as a tool for reflection and as an internal guide to behavior is poorly developed. Similarly, the present view proposes that more advanced reasoning begins when individuals ask questions, not of others, but of themselves, and, through the gradual internalization of the linguistic elements and the pattern of hypothetico-predictive argumentation, develop the ability to 'talk to themselves', which constitutes the essence of reflective thought and allows one to internally test alternative propositions (e.g. nominal, categorical, causal, and theoretical) to arrive at internally reasoned decisions.

The use of hypothetico-predictive argumentation in concept construction

Newton et al. (1999) claimed that student argumentation facilitates clearer conceptual understanding. As they put it:

> Talking offers an opportunity for conjecture, argument and challenge. In talking, learners will articulate reasons for supporting particular conceptual understandings and attempt to justify their views. Others will challenge, express doubts and present alternatives, so that a clearer conceptual understanding will emerge. (1999: 554)

Several previous researchers have advanced the view that argument not only helps in the clarification of concepts, but that the development of procedural (i.e. operational) knowledge is necessary for the construction of concepts and conceptual systems (for example, Anderson 1980, Fosnot 1996, Inhelder and Piaget 1958, Karplus 1977, Kuhn et al. 1988, Piaget 1964, von Glasersfeld 1995).

Let us briefly consider how hypothetico-predictive argumentation may be involved in concept construction and conceptual change. Take, for example, the biological question of the origin of present-day life forms: What caused the diversity of living things presently found on Earth? Of course several theories have been advanced to answer this question, two of the more prominent ones being the biblically inspired theory of special creation and Darwin's theory of evolution. These theories represent competing conceptual systems. How does one decide which conceptual system to accept and which to reject? The present claim is that a thoughtful decision requires the generation of several hypothetico-predictive arguments. The arguments can occur during a debate/discussion among others, or they can occur internally, as an internal dialog. Here, for example, are two of several relevant hypothetico-predictive arguments:

> *If* . . . present-day life forms were created by God, 'each according to its own kind', as stated in the book of Genesis [special creation theory],
> *and* . . . present-day life forms in several locations are compared [planned test],
> *then* . . . each life form should fit into one or another category/species, depending on its initially created form [prediction].
> *But* . . . intermediate life forms do exist.
> *Therefore* . . . special creation theory is not supported [conclusion].

Alternatively:

> *if* . . . present-day life forms arose through evolution [evolution theory],
> *and* . . . present-day life forms in several locations are compared [planned test],
> *then* . . . intermediate life forms, between two categories/species, should exist [prediction].
> *And* . . . intermediate life forms do exist [observed result].
> *Therefore* . . . evolution theory is supported [conclusion].

In theory, engaging in such arguments is necessary to change from a creationist to an evolutionist position (i.e. to undergo conceptual change). Thus, students lacking facility with stage 5 reasoning skills should have difficulty undergoing conceptual change and rejecting scientific misconceptions such as special creation (for example, Lawson and Weser 1990, Lawson and Worsnop 1992, Lawson et al. 1993, Lawson et al. 2000a,b). However, the claim is not that conceptual change occurs easily and quickly. Indeed, in Darwin's case it took several years to change from the creationist view and become an evolutionist (Gruber and Barrett 1974). There appear to be several reasons why conceptual change can be a lengthy process, in addition to lack of reasoning skill. These include lack of understanding of the new conception and lack of sufficient evidence that the old conception is somehow 'incorrect' and the new conception is an improvement (for example, Lawson 2003, Posner et al. 1982, Strike and Posner 1992).

The use of hypothetico-predictive argumentation in the science classroom

Let us now turn to a discussion of how students can be encouraged to construct and use hypothetico-predictive arguments in the classroom. Hypothetico-predictive

Figure 4. This is what happens when an inverted cylinder is placed over a burning candle sitting in a pan of water.

argumentation can be expected to achieve two instructional goals. First, given sufficient time and sufficient reasoning skill and reflectivity on the students' part, it should lead to conceptual acquisition and to conceptual change. And second, it should lead to the development of increased consciousness and skill in constructing and using such arguments (i.e. in intellectual development from one stage to the next). A well-known classroom investigation can exemplify these points. The investigation provides an opportunity for students to generate and test alternative hypotheses about the cause of water rise in an inverted cylinder (Elementary Science Study 1974, Lawson 1999, Lawson et al. 2000c). Primary concepts acquired are those embedded in kinetic molecular theory (e.g. atoms, molecules, kinetic energy, energy transfer, air pressure), while conceptual change centers around two primary scientific misconceptions (i.e. flames 'consume' oxygen molecules and a pulling force called suction exists).

To initiate the investigation, students place an inverted cylinder over a burning candle sitting upright in a shallow pan of water. The candle is held upright by a small piece of modeling clay (see figure 4). Shortly after the cylinder is placed over the burning candle, the flame goes out and water rushes up into the cylinder.

Generating causal questions and alternative hypotheses

These puzzling observations raise two key causal questions: Why does the flame go out? And why does the water rise? The first causal question is difficult to investigate, but the second easily lends itself to student hypothesis generation and test. Student proposed hypotheses (claims) generated via group brainstorming and the use of analogical reasoning often include the following:

1. The oxygen is burned up (consumed) creating a partial vacuum. Thus, the water is 'sucked' up into the cylinder to fill the partial vacuum.
2. As the candle burns, it converts oxygen gas (O_2 molecules) into carbon dioxide gas (CO_2 molecules). The carbon dioxide molecules dissolve in the water more readily than the original oxygen molecules, thus a partial vacuum is created. The water then rises to fill the partial vacuum.
3. The candle's heat causes the air around it to expand and/or escape from the open end of the cylinder. After the candle goes out, the air becomes cooler. Cooling reduces the internal air pressure. Thus, water is pushed into the cylinder by greater external air pressure.

Constructing arguments to test the consumed-oxygen hypothesis

Now that students have generated some alternative hypotheses, the next task is to gather evidence to put them to the test. The following attempts to link hypothesis 1 with a specific prediction that can be used in its test:

> *If* . . . hypothesis 1 is correct, that is if water rises in the cylinder because oxygen is consumed creating a partial vacuum [consumed-oxygen hypothesis],
> *and* . . .the height of water rise with one, two, three, or more candles is measured [planned test],
> *then* . . .the height of water rise *should* be the same regardless of the number of burning candles [prediction].

The height of water rise should be the same presumably because there is only so much oxygen in the cylinder to be burned. Thus, more candles will burn the available oxygen faster than fewer candles, but the additional candles will not burn more oxygen. Hence, the water should raise the same. These statements represent a *warrant* because they function to link the hypothesis being tested to a prediction. In other words, they function to convince others that the prediction does in fact follow from the hypothesis and the planned test.

Interestingly, on several occasions when a student argues that the consumed-oxygen hypothesis leads to the above prediction (based on the stated warrant), one or more other students voice skepticism. In other words, they question whether the prediction does in fact follow from the hypothesis and planned test. Instead, they suggest that perhaps more candles do in fact burn more oxygen because the additional candles cause a greater mixing of the available oxygen, hence cause more oxygen to reach the flames and burn. This counter-argument now necessitates the construction of *backings* for the original warrant (e.g. perhaps more candles do create more mixing, thus *slightly* more oxygen is burned, but *surely* not enough more to account for the much greater water rise with three candles than with one).

Once the planned test is conducted and once students observe that additional candles do in fact cause considerably more water rise, most begin to doubt the veracity of the consumed-oxygen hypothesis. The consumed-oxygen hypothesis also leads the prediction that the water should rise *while* the candles are burning, not after they have gone out. Therefore, the observation that most of the water rising occurs after the candles go out casts further doubt on the consumed-oxygen hypothesis. One complete hypothetico-predictive argument used to test the consumed-oxygen hypothesis can be summarized as follows:

> *If* . . . water rises because oxygen is consumed creating a partial vacuum [consumed-oxygen hypothesis],
> *and* . . . the height of water rise with one, two, three, or more candles is measured [planned test],
> *then* . . . the height of water rise should be the same regardless of the number of burning candles [prediction].
> *But* . . . the water rises considerably higher with additional burning candles [observed result].
> *Therefore* . . . the consumed-oxygen hypothesis is not supported [conclusion].

Testing the dissolving-CO_2 hypothesis

Does the same form of hypothetico-predictive argumentation apply to testing hypothesis 2? Hypothesis 2 claims that the water rises due to dissolving carbon dioxide molecules and a resulting decrease in internal pressure. Students can test this dissolving-CO_2 hypothesis by comparing the height of water rise in a pan with CO_2-saturated water with one containing normal water (all other thing being equal, i.e. method of differences). Dry ice or sodium bicarbonate and acid can be used to produce excess CO_2. The planned test along with the dissolving-CO_2 hypothesis lead to the prediction that the water level should raise less in the cylinder placed in the CO_2 saturated water than in the one placed in normal water:

If . . . the water rises because carbon dioxide dissolves rapidly into the water [dissolving-CO_2 hypothesis],
and . . . the height of water rise in two containers is compared – one with CO_2 saturated water and one with normal water [planned test],
then . . . the water should rise less in the container with the CO_2 saturated water than in the container with the normal water [prediction].

The warrant needed to link the dissolving-CO_2 hypothesis to the prediction goes something like this. Dissolving CO_2 molecules presumably cause a reduction of air pressure in the cylinder. This reduction in turn causes the water rise. Consequently, when the water is already saturated with CO_2 molecules, the newly created CO_2 molecules cannot escape into the water; hence, the internal pressure will not be reduced and the water will not rise. But it turns out that when the planned test is conducted, the water level raises the same in both containers. Therefore, the dissolving-CO_2 hypothesis is not supported:

If . . . the hypothesis is correct [dissolving-CO_2 hypothesis],
and . . . the height of water rise in the two containers is compared [planned test],
then . . . the water should rise less in the container with the CO_2 saturated water than in the container with the normal water [prediction].
But . . . the water rises the same in both containers [observed result].
Therefore . . . the dissolving-CO_2 hypothesis is not supported [conclusion].

Testing the air-expansion hypothesis

Hypothesis 3, the air-expansion hypothesis, leads to at least three predictions: (1) bubbles should be seen escaping from the bottom of the cylinder (*warrant*: assuming that the cylinder is quickly placed over the candles while the air is still expanding, the expanding air will escape from the bottom); (2) more candles should cause more water to rise (*warrant*: more candles will heat more air; thus more will escape, which in turn will be replaced by more water; although one candle burning over a longer time period releases as much energy as three candles burning a shorter time, one candle will not raise the cylinder's air temperature as much because energy is dissipated rather quickly); (3) the water should continue to rise *after* the candles have gone out (*warrant*: air cools after the candles go out because the air molecules inside the cylinder strike the cylinder's surface and transfer kinetic energy (motion) to the cylinder; thus they lose speed, resulting in lower air pressure on the water's surface). Each predicted result occurs. Therefore, the air-expansion hypothesis is supported:

If . . . water rises due to expansion, escape and cooling of air [air-expansion hypothesis],
and . . . we observe the bottom of the cylinder while it is placed over the burning candles into the water [planned test],
then . . . bubbles should be observed coming from the bottom of the cylinder [prediction].
And . . . bubbles are observed coming from the bottom of the cylinder [observed result].
Therefore . . . the air-expansion hypothesis is supported [conclusion].

Now that student research has produced evidence for one of the alternative hypotheses (the air-expansion hypothesis) and against the alternatives, the next step is to write a laboratory report presenting the hypothetico-predictive arguments. The goal of the laboratory report is to convince the reader that the air-expansion hypothesis is 'correct' *and* that the other hypotheses are 'incorrect'. Of course, this statement in no way implies that hypothetico-predictive arguments can prove or disprove hypotheses in any ultimate sense. Hypotheses cannot be proved because any number of hypotheses may lead to the same prediction. Hypotheses cannot be disproved because a mismatch of predictions and observed results may arise from a poor test, rather than from a faulty hypothesis (i.e. Lawson 1993a). Statisticians often refer to the mistaken rejection of a 'true' hypothesis as a type I error and to the mistaken acceptance of a 'false' hypothesis as a type II error (for example, Glass and Stanley 1970). The appendix discusses this issue further with a look at the relationship between hypothetico-predictive argumentation and conditional logic.

Additional hypothetico-predictive arguments embedded in biology instruction

During the past several semesters, we have been teaching an introductory college-level biology course in which laboratory and field investigations engage students in constructing and using several additional hypothetico-predictive arguments. The arguments are presented both verbally and in writing so that they can be evaluated by other students and by the instructor. To provide a sense of the scope and diversity of the arguments, table 1 presents some of the arguments that arise during the semester. As you can see, the arguments for investigations 1 and 5 use circumstantial evidence and sign arguments (i.e. arguments based on relationships of objects, events or situations to their characteristics). Investigation 7 uses correlational evidence and concomitant variation arguments (i.e. arguments about possible cause–effect relationships by the evaluation of the degree of correlation between two variables). Investigations 2, 3, 4, 6, and 8 use experimental evidence and method of differences arguments (i.e. arguments that establish a cause in that it alone varies when the effect varies).

The basic instructional approach employed during the investigations is to provide students with inquiry-based experiences such as the candle burning investigation. The investigations allow students to generate and test alternative hypotheses, and verbally present arguments for and against the alternatives during classroom discussions. Once arguments and counter-arguments have been presented and discussed, students are asked to construct written arguments by using the words *If/and/then/Therefore* and by filling in the 'clouds' and boxes shown in figure 2 with the key elements of their arguments. In theory, this emphasis on verbal and then on written arguments should encourage internalization and the movement from stage 3, to stage 4, and then possibly to stage 5 reasoning.

Conclusion

In conclusion, a pattern of scientific argumentation has been presented in which scientists respond to causal questions with the generation and test of alternative hypotheses thorough cycles of hypothetico-predictive argumentation. In theory, hypothetico-predictive arguments are employed to test causal claims that exist on at least two levels (designated stage 4 in which the causal claims are perceptible, and

Table 1. Example hypothetico-predictive arguments constructed during some of the laboratory and field investigations.

Investigation 1. What can be inferred from animal structure?
If . . . the sharp canine teeth found in this animal skull were used to capture prey,
and . . . the position of the skull's eye sockets is noted,
then . . . the eye sockets should be pointed forward (presumably as needed by prey-capturing animals for the good depth perception).

Investigation 2. How smart are animals?
If . . . isopods congregate at the left end of the trough because they followed a leader isopod,
and . . . isopods are placed in the trough as a group versus placed in one at a time,
then . . . the isopods should congregate at the left end only when they are placed in as a group (i.e. presumably when there is a leader).

Investigation 3. What causes intra-specific variation?
If . . . the observed 'bell-shaped' distribution of snail shell lengths is caused by environmental factors (i.e., by wave action and by predators),
and . . . snails are reared for several generations in an environment without waves and predators,
then . . . the resulting distribution of shell lengths should not be 'bell-shaped'.

Investigation 4. What determines specific characteristics in fruit flies?
If . . . the observed difference in fly eye color (i.e. red vs. brown eyes) is caused by dominant and recessive genes,
and . . . red-eyed flies are mated with brown-eyed flies,
then . . . the eye colors of their offspring should occur in a 3 to 1 ratio.

Investigation 5. What do fossils tell us about life in the past?
If . . . all living things were created during a short period of time by an act of God,
and . . . the fossils found in different rock layers of the Grand Canyon are compared,
then . . . fossils found in the rock layer just above the layer corresponding to the time of creation should contain virtually all fossil types.

Investigation 6. How do species evolve?
If . . . the peppered moth population changed from predominately white to black due to natural selection,
and . . . an equal number of white and black moths are placed on light and dark trees,
then . . . birds should capture more white moths on dark trees and more black moths on light trees.

Investigation 7. Which human characteristics co-vary?
If . . . breast implants cause connective tissue disease,
and . . . disease incidence in women with implants is compared to that in women without implants,
then . . . disease incidence should be higher in the implant group than in the non-implant group.

Investigation 8. Why are plants green?
If . . . plants are green because they reflect green light,
and . . . water plants are placed under green light and then under red light,
then . . . more oxygen bubbles should escape from the plants when under red light (presumably green light will be reflected but red light will be absorbed, hence will drive photosynthesis and produce the oxygen bubbles).

stage 5 in which the causal claims are imperceptible). Origins of the ability to construct and comprehend hypothetico-predictive arguments at the highest level can be traced to pre-verbal reasoning of the sensory-motor child and the gradual internalization of verbally mediated arguments involving nominal, categorical, causal and, finally, theoretical propositions. Presumably, the ability to construct and comprehend hypothetico-predictive arguments (an aspect of procedural knowledge) is necessary for the construction of conceptual knowledge (an aspect of declarative knowledge) because such arguments are used during concept construction and conceptual change. Science instruction that focuses on the generation and debate of hypothetico-predictive arguments should improve students' argumentative/reasoning skills. Such improvements should be linked to the instructor's ability to engage students in verbal and written discourse, and to reflect on the arguments and science concepts being taught.

Acknowledgements

This material is based upon research partially supported by the National Science Foundation under grant number DUE 0084434. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Anderson, J. R. (1980). *Cognitive psychology and its implications* (San Francisco: Freeman).

Biela, A. (1993). *Psychology of analogical inference* (Stuttgart: S. Hirzel Verlag).

Boden, M. A. (1994). What is creativity? In M. A. Boden (Ed.), *Dimensions of creativity* (Cambridge, MA: MIT Press).

Bruner, J. (1962). *On knowing: essays for the left hand* (Cambridge, MA: Harvard University Press).

Chamberlain, T. C. (1965). The method of multiple working hypotheses. *Science*, 148, 754–759. (originally published in 1898).

Conway, D. A. and Munson, R. (1990). *The elements of reasoning* (Belmont, CA: Wadsworth).

Diamond, A. (1990). The development and neural bases of inhibitory control in reaching in human infants and infant monkeys. In A. Diamond (Ed.), *The development and neural basis of higher cognitive functions* (New York: Academy of Sciences).

Dempster, F. N. (1992). Resistance to interference: developmental changes in a basic processing mechanism. *Developmental Review*, 12, 45–57.

Dreistadt, R. (1968). An analysis of the use of analogies and metaphors in science. *Journal of Psychology*, 68, 97–116.

Elementary Science Study (1974). *Teachers' guide for attribute games and problems* (New York: McGraw-Hill).

Finke, R. A., Ward, T. B. and Smith, S. M. (1992). *Creative cognition: theory research and practice* (Cambridge, MA: The MIT Press).

Fosnot, C. T. (1996). Constructivism: a psychological theory of learning. In C. T. Fosnot (Ed.), *Constructivism: theory, perspectives, and practice* (New York: Teacher's College Press).

Freeley, A. J. (1976). *Argumentation and debate* (Belmont, CA: Wadsworth).

Gentner, D. (1989). The mechanisms of analogical learning. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (Cambridge: Cambridge University Press).

Glass, G. V. and Stanley, J. C. (1970). *Statistical methods in education and psychology* (Englewood Cliffs, NJ: Prentice-Hall).

Gruber, H. E,, and Barrett, P. H. (1974). *Darwin on man: a psychological study of scientific creativity* (New York: Dutton).

Hempel, C. (1966). *Philosophy of natural science* (Upper Saddle River, NJ: Prentice-Hall).

Hestenes, D. (1992). Modeling Games in a Newtonian World. *American Journal of Physics*, 55, 440–454.

Hiatt, D. A. (1975). *True, false or in between* (Lexington, MA: Ginn).

Hoffman, R. R. (1980). Metaphor in science. In P. R. Honeck, & R. R. Hoffman (Eds.), *The psycholinguistics of figurative language* (Hillsdale, NJ: Erlbaum).

HOFSTADTER, D. R. (1981). Metamagical themas: how might analogy, the core of human thinking, be understood by computers? *Scientific American*, 249, 18–29.

HOLLAND, J. H., HOLYOAK, K. J., NISBETT, R. E. and THAGARD, P. R. (1986). *Induction: processes of inference, learning, and discovery* (Cambridge, MA: The MIT Press).

INHELDER, B. and PIAGET, J. (1958). *The growth of logical thinking from childhood to adolescence* (New York: Basic Books).

JOHNSON, M. (1987). *The body in the mind* (Chicago, IL: University of Chicago Press).

KARPLUS, R. (1977). Science teaching and the development of reasoning. *Journal of Research in Science Teaching*, 14, 169–175.

KOESTLER, A. (1964). *The act of creation* (London: Hutchinson).

KUHN, D., AMSEL, E. and O'LOUGHLIN, M. (1988). *The development of scientific thinking skills* (San Diego: Academic Press).

LAWSON, A. E. (1993a). On why falsifiability does not exist in theory testing: a reply to Smith and Siegel. *Journal of Research in Science Teaching*, 30, 603–606.

LAWSON, A. E. (1993b). Deductive reasoning, brain maturation, and science concept acquisition: Are they linked? *Journal of Research in Science Teaching*, 30, 1029–1051.

LAWSON, A. E. (1999). What should students learn about the nature of science and how should we teach it? *Journal of College Science Teaching*, 28, 401–411.

LAWSON, A. E. (2000a). The generality of hypothetico-deductive reasoning: making scientific thinking explicit. *The American Biology Teacher*, 62, 482–495.

LAWSON, A. E. (2000b). How do humans acquire knowledge? And what does that imply about the nature of knowledge? *Science and Education*, 9, 577–598.

LAWSON, A. E. (2003). *The neurological basis of learning, development and discovery: implications for science and mathematics instruction* (Dordrecht: Kluwer).

LAWSON, A. E., BAKER, W. P., DIDONATO, L., VERDI, M. P. and JOHNSON, M. A. (1993). The role of physical analogues of molecular interactions and hypothetico-deductive reasoning in conceptual change. *Journal of Research in Science Teaching*, 30, 1073–1086.

LAWSON, A. E., CLARK, B., CRAMER-MELDRUM, E., FALCONER, K. A., SEQUIST, J. M. and KWON, Y. J. (2000a). The development of scientific reasoning in college biology: do two levels of general hypothesis-testing skills exist? *Journal of Research in Science Teaching*, 37, 81–101.

LAWSON, A. E., DRAKE, N., JOHNSON, J., KWON, Y. J. and SCARPONE, C. (2000b). How good are students at testing alternative explanations involving unseen entities? *The American Biology Teacher*, 62, 246–252.

LAWSON, A. E. and KRAL, E. (1985). Developing formal reasoning through the study of English. *The Educational Forum*, 49, 211–226.

LAWSON, A. E., LEWIS, C. M. and BIRK, J. P. (2000c). Why do students 'cook' data? *Journal of College Science Teaching*, 29, 191–198.

LAWSON, A. E. and WESER, J. (1990). The rejection of nonscientific beliefs about life: effects of instruction and reasoning skills. *Journal of Research in Science Teaching*, 27, 589–606.

LAWSON, A. E. and WORSNOP, W. A. (1992). Learning about evolution and rejecting a belief in special creation: effects of reflective reasoning skill, prior knowledge, prior beliefs and religious commitment. *Journal of Research in Science Teaching*, 29, 143–166.

LEVINE, D. S. and PRUEITT, P. S. (1989). Modeling some effects of frontal lobe damage: novelty and perseveration. *Neural Networks, 2*, 103–116.

LEWIS, R. W. (1988). Biology: a hypothetico-deductive science. *The American Biology Teacher*, 54, 137–152.

LURIA, A. R. (1991). *The role of speech in the regulation of normal and abnormal behavior* (Oxford: Pergamon).

MEDAWAR, P. B. (1969). *Induction and intuition in scientific thought* (Philadelphia, PA: American Philosophical Society).

MELTZOFF, A. N. (1990). Towards a developmental cognitive science. In A. Diamond (Ed.), *The development and neural basis of higher cognitive functions* (New York: Academy of Sciences).

NEWTON, P., DRIVER, R. and OSBORNE, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, 21, 553–576.

OLSON, R. C. (1969). *Meaning and argument: elements of logic* (New York: Harcourt Brace and World).

PIAGET, J. (1928). *Judgment and reasoning in the child* (Paterson, NJ: Littlefield Adams).

PIAGET, J. (1954). *The construction of reality in the child* (New York: Basic Books).

Piaget, J. (1964). Cognitive development in children: development and learning. *Journal of Research in Science Teaching*, 2, 176–186.

Piaget, J. (1976). *The grasp of consciousness* (Cambridge: Harvard University Press).

Piaget, J. and Inhelder, B. (1969). *The psychology of the child* (New York: Basic Books).

Platt, J. R. (1964). Strong inference. *Science*, 146, 347–353.

Popper, K. (1965). *Conjectures and refutations. the growth of scientific knowledge* (New York: Basic Books).

Popper, K. R. (1959). *The logic of scientific discovery* (New York: Basic Books).

Posner, G., Strike, K., Hewson, P. and Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, 66, 211–227.

Salmon, M. H. (1995). *Introduction to logic and critical thinking* (Fort Worth, TX: Harcourt Brace).

Sherman, P. W. and Flaxman, S. M. (2001). Protecting ourselves from food. *American Scientist*, 89, 142–151.

Sternberg, R. J. and Davidson, J. E. (Eds.) (1995). *The nature of insight* (Cambridge, MA: The MIT Press).

Strike, K. and Posner, G. (1992). A revisionist theory of conceptual change. In R. Duschl, & R. Hamilton (Eds.), *Philosophy of science, cognitive psychology, and educational theory and practice* (Albany, NY: State University of New York), 147–176.

Toulmin, S. E. (1958). *The uses of argument* (Cambridge: Cambridge University Press).

Toulmin, S., Rieke, R. and Janik, A. (1984). *An introduction to reasoning*, 2nd ed. (New York: Macmillan).

Von Glasersfeld, E. (1995). *Radical constructivism: a way of knowing and learning* (London: Falmer Press).

Vygotsky, L. S. (1962). *Thought and language* (Cambridge, MA: MIT Press).

Wandersee, J. H., Mintzes, J. J. and Novak, J. D. (1994). Research on alternative conceptions in science. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (New York: Macmillan).

Warnick, B. and Inch, E. S. (1989). *Critical thinking and communication: the use of reason in argument* (New York: Macmillan).

Wong, E. D. (1993). Self-generated analogies as a tool for constructing and evaluating explanations of scientific phenomena. *Journal of Research in Science Teaching*, 30, 367–380.

Appendix: the logic of hypothetico-predictive reasoning

How does hypothetico-predictive argumentation relate to standard rules of conditional logic such as *modus tollens* and *modus ponens*? The logic of *modus tollens* reads as follows: p implies q, not-q, therefore, not-p. For hypothesis 1 in the candle-burning context (the consumed-oxygen hypothesis), we get the following:

If . . . water rises because oxygen is consumed [p],
and . . . the height of water rise with additional burning candles is measured,
then . . . the height should be the same regardless of the number of burning candles [q].
But . . . the water rises higher with additional burning candles [not-q].
Therefore . . . the consumed-oxygen hypothesis is not correct [not-p].

However, as previously pointed out, the failure of an observed result to match a predicted result may stem from a faulty test, rather than a faulty hypothesis. Consequently, a more reasonable application of *modus tollens* might read as follows:

If . . . water rises because oxygen is consumed [p],
and . . . the height of water rise with additional burning candles is measured,
then . . . the height should be the same regardless of the number of burning candles [q] – assuming nothing goes wrong with the test – the test was perfectly controlled.
But . . . the water rises higher with additional burning candles [not-q].

Therefore . . . most likely the consumed-oxygen hypothesis is not correct [not-p] – unless something did go wrong with the test!

Now consider the logic of *modus ponens*: p implies q, p, therefore q. Interestingly, *modus ponens* does not appear to apply, as the following illustrates:

If . . . water rises because oxygen is consumed [p],
and . . . the height of water rise with additional burning candles is measured,
then . . . the height should be the same regardless of the number of burning candles [q].
And . . . water rises because oxygen is consumed [p].
Therefore . . . the height of water rise should be the same regardless of the number of burning candles [q].

Clearly, this argument makes no sense. The point of hypothetico-predictive reasoning is to test an idea. On the other hand, the point of *modus ponens* is to generate a 'logical' prediction. So, once again, a standard logical rule seems to fail to capture the essence of hypothetico-predictive reasoning. Interestingly, the logical fallacy known as affirming the consequent seems to do a better job (cf. Hempel 1966: 6–7). Affirming the consequent reads as follows: p implies q, q, therefore p. In the candle-burning situation we get the following:

If . . . water rises due to expansion, escape and cooling of air [p],
and . . . we observe the cylinder's bottom while placed over the burning candles into the water [planned test],
then . . . bubbles should be observed coming from the bottom [q].
And . . . bubbles are observed coming from the bottom [q].
Therefore . . . water rises due to expansion, escape and cooling of air [p].

But as previously noted, drawing this conclusion represents a logical fallacy. The conclusion is also 'unreasonable' because the bubbles might occur for other reasons (i.e. alternative hypotheses exist that have not been tested and rejected). For example, perhaps the bubbles were simply caused by the unheated air initially in the cylinder. Consequently, the more reasonable conclusion is that the initial hypothesis has been supported, but one cannot be certain that it is correct. The following summarizes the necessary modifications. For *modus tollens*:

If . . . p,
and . . . the planned test,
then . . . *probably* q [assuming that nothing goes wrong with the test].
But . . . not-q.
Therefore . . . *probably* not-p [meaning that the hypothesis p is not supported, but is not disproved].

And for affirming the consequent:

If . . . p,
and . . . the planned test,
then . . . *probably* q [assuming that nothing goes wrong with the test].
And . . . q.
Therefore . . . *possibly* p [meaning that the hypothesis is supported, but not proven as other hypotheses could lead to the same prediction].

Consequently, the essence of hypothetico-predictive reasoning is not fully captured by the logic of *modus tollens* and *modus ponens*. But this does not mean that humans are unreasonable. Said another way, it would appear that our minds do not necessarily reason with these rules of conditional logic.