ANTONE E. LAWSON

THE NATURE AND DEVELOPMENT OF SCIENTIFIC REASONING: A SYNTHETIC VIEW

ABSTRACT. This paper presents a synthesis of what is currently known about the nature and development of scientific reasoning and why it plays a central role in acquiring scientific literacy. Science is viewed as a hypothetico-deductive (HD) enterprise engaging in the generation and test of alternative explanations. Explanation generation and test requires the use of several key reasoning patterns and sub-patterns. Reasoning at the highest level is complicated by the fact that scientific explanations generally involve the postulation of non-perceptible entities, thus arguments used in their test require sub-arguments to link the postulate under test with its deduced consequence. Science is HD in nature because this is how the brain spontaneously processes information whether it basic visual recognition, every-day descriptive and causal hypothesis testing, or advanced theory testing. The key point in terms of complex HD arguments is that if sufficient chunking of concepts and/or reasoning sub-patterns have not occurred, then one's attempt to construct and maintain such arguments in working memory and use them to draw conclusions and construct concepts will "fall apart." Thus, the conclusions and concepts will be "lost." Consequently, teachers must know what students bring with them in terms of their stages of intellectual development (i.e., preoperational, concrete, formal, or post-formal) and subject-specific declarative knowledge. Effective instruction mirrors the practice of science where students confront puzzling observations and then personally participate in the explanation generation and testing process – a process in which some of their ideas are contradicted by the evidence and by the arguments of others.

KEY WORDS: brain functioning, intellectual development, instruction, reasoning, scientific literacy

In some countries, the development of the ability to reason scientifically has long been a central goal of education in general and of science and mathematics education in particular. In the United States, for example, the Educational Policies Commission (1961) advanced the proposition that the central goal of American education is the development of students' rational powers, which in their words constitute "the essence of the ability to think" (p. 5). In a subsequent document (Educational Policies Commission, 1966), the commission identified science and mathematics education as key vehicles to advance that central goal. Again in their words: "What is being advocated here is not the production of more physicists, biologists, or mathematicians, but rather the development of persons whose approach

International Journal of Science and Mathematics Education (2004) 2: 307–338

to life as a whole is that of a person who thinks – a rational person" (p. 16). As reviewed by Hand, Prain and Yore (2001) current international reform documents echo the same theme and argue that scientific reasoning abilities and habits of mind lie at the heart of scientific literacy, which involves: (1) the abilities and habits of mind to construct understanding, (2) understanding the central concepts and unifying theories of science, and (3) the ability to communicate to inform and persuade others to take action related to those concepts and theories.

What do we now know about the nature and development of rational (i.e., scientific) reasoning abilities? Have we learned enough to better instruct students in ways that will help them become better reasoners in a general sense and become scientifically literate? This paper will argue that the answer is yes. The argument will be developed in steps. We begin by explicating the nature of scientific reasoning through a case study. We then discuss the neurological basis of such reasoning, trace its course and causes of development during childhood, adolescence, and early adulthood, identify its relationship to the acquisition of science concepts and to students' awareness of the nature of science, and conclude by discussing instructional methods that have been found to promote the development of scientific reasoning abilities and scientific literacy.

A Few Key Definitions and Clarifications

Before introducing the case study, a few definitions and clarifications are in order. A reasoning pattern is defined as a mental strategy, plan, or rule used to process information and derive conclusions that go beyond direct experience. As such, reasoning patterns are part of one's procedural or operative knowledge – one's "how to" knowledge – as opposed to one's figurative or declarative knowledge – one's "that is" knowledge (e.g., Piaget, 1970; Anderson, 1980). Procedural knowledge, which is expressed through performance, is often implicit in the sense that we may not be conscious that we have it or precisely when or how it was acquired. The word "development" is often used in conjunction with the acquisition of procedural knowledge. On the other hand, declarative knowledge is explicit – that is we often know that we have it and when and how it was acquired. The word "learning" is often used in conjunction with the acquisition of declarative knowledge. As will be argued, scientific reasoning consists of an overall pattern of reasoning, which can be characterized as hypothetico-deductive, as well as several sub-patterns. Inhelder and Piaget (1958) referred to these sub-patterns as formal operational schemata (e.g., combinatorials, proportions, correlations). Logicians often refer to them as "methods," or "forms" of argumentation such as argument by analogy, method of difference, method of agreement, and concomitant variation (e.g., Tidman & Kahane, 2003; Warnick & Inch, 1989). Recent neurological research indicates that the procedural knowledge patterns, once acquired, reside in neural networks that are hierarchical in nature. The hierarchical networks culminate in single neurons (see later discussion of chunking) located in the brain's prefrontal cortex (Wallis, Anderson & Miller, 2001). Alternatively, declarative knowledge resides in associative memory, which is located primarily in the hippocampus, the limbic thalamus and the basal forebrain (Kosslyn & Koenig, 1995). Further, it appears that the conscious recollection of procedural knowledge is independent of the medial temporal lobe, thus depends on other brain systems such as the neo-striatum while the storage and recollection of declarative knowledge depends on the functional integrity of the medial temporal lobe (Squire & Zola-Morgan, 1991).

Importantly, there appears to be two ways to acquire both procedural and declarative knowledge – to get new information into long-term memory. One way is through shear repetition and/or via emotionally charged contexts. Repetition and emotion can "burn" new input into one's synapses essentially by boosting pre-synaptic activity to a high enough level to create new functional synaptic connections (e.g., Grossberg, 1982). Students can memorize their multiplication tables and the positions of letters on a keyboard in this "rote" way. They can also learn to solve proportions problems in a rote way by use of a "cross-multiplication" algorithm (e.g., $4/6 = 6/X$, $(4)(X) = (6)(6)$, $(4)(X) = 36$, $X = 36/4$, $X = 9$). Unfortunately, in spite of the fact that students can cross multiply and "solve" such problems, they typically have no idea why the algorithm works or how to solve "real" problems involving proportional relationships. For example, most 12 year olds in the United States can easily tell you that X equals nine in the previous equation, but when given the "Cylinders" problem shown in Figure 1, they incorrectly predict that water will rise to the 8th mark "... because it rose two more before, from 4 to 6, so it will rise 2 more again, from 6 to 8."

Fortunately, there is a second way to get information into long-term memory. That way is to form new functional synaptic connections by linking new input with prior ideas (Grossberg, 1982). When neural activity is simultaneously boosted by new input and by prior ideas, the resulting pre- and post-synaptic activities combine to create new functional connections. This connectionist (or constructivist) way of learning has several advantages, not the least of which is that learning is not rote in the sense that it is connected to what you already know, thus becomes much more useful in reasoning and problem solving. In the case of proportions this means that students not only know how to solve for X, they also know when to use a

To the right are drawings of a wide and a narrow cylinder. The cylinders have equally spaced marks on them. Water is poured into the wide cylinder up to the 4th mark (see A). This water rises to the 6th mark when poured into the narrow cylinder (see B). Both cylinders are emptied, and water is poured into the wide cylinder up to the 6th mark. How high will this water rise when poured into the empty narrow cylinder?

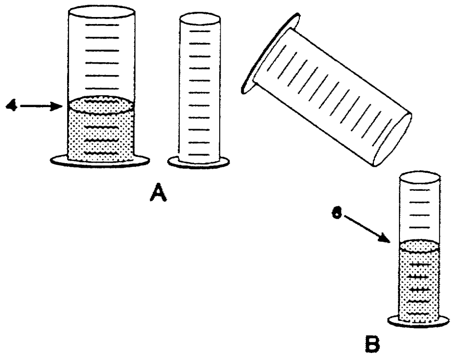


Figure 1. The cylinders problem.

proportions strategy and when not to, i.e., they know when other strategies, such as addition and subtraction, should be used instead. The point is that if we want students to become good problem solvers and good scientific thinkers, we cannot teach in ways that lead to rote learning. Instead, we need to become connectionist teachers. With this in mind, let's turn to the case study of scientific reasoning.

A Case Study: Indentifying Scientific Reasoning Patterns

Silver salmon are born in the cool, quiet headwaters of freshwater streams in the Pacific Northwest. Young salmon swim downstream to the Pacific Ocean where they grow and mature sexually. They then return to the freshwater streams and swim upstream to ultimately lay eggs or deposit sperm in the headwaters before dying. By tagging young salmon, biologists discovered that mature salmon actually return to reproduce in precisely the same headwaters where they were born some years earlier. This discovery raised a very interesting causal question: how do returning salmon find their home stream? In other words, what causes them to end up in their home stream?

During the 1960s, biologist A.D. Hasler proposed a number of alternative explanations – alternative causal hypotheses. For instance, people often navigate by sight. Perhaps salmon do as well. Returning salmon may recall objects, such as large rocks, they saw while swimming downstream on their way to the ocean. Studies of migratory animals suggested alternative explanations. For example, biologists knew that migratory eels are enormously sensitive to dissolved chemicals in water. Perhaps salmon are as well. Biologists also knew that homing pigeons navigate using the

Earth's magnetic field. Perhaps salmon are also sensitive to the magnetic field and use it to find their home stream. Thus, by borrowing explanations from other possibly similar contexts as tentative explanations in the present context – by using analogies – Hasler generated three alternative hypotheses for salmon navigation: (1) salmon use sight; (2) salmon smell chemicals specific to their home stream; and (3) salmon use the Earth's magnetic field.

The use of analogies in this way, sometimes called abduction, analogical transfer or analogical reasoning, to generate possible explanations (causal hypotheses) is a very general and creative process (e.g., Biela, 1993; Coll & Treagust, 2002; Dreistadt, 1968; Finke, Ward & Smith, 1992; Gentner, 1989; Hestenes, 1992; Hoffman, 1980; Hofstadter, 1981; Holland, Holyoak, Nisbett & Thagard, 1986; Johnson, 1987; Koestler, 1964). In the words of philosopher Charles Peirce (as quoted in Hanson, 1958, p. 85), "All ideas of science come to it by way of Abduction." By abduction Peirce means that scientists generate new explanations by abducting/stealing explanations from old contexts (based on contextual similarities) and attempt to use them as possible explanations in new contexts. This means that one's store of declarative knowledge is the source of hypotheses. Of course in addition to the three explanations listed above, other possibilities remain. Indeed none of the three may be correct. Or perhaps salmon use two of the three methods, or perhaps all three. By generating all possible combinations of explanations (i.e., by using combinatorial reasoning – the reasoning pattern used to systematically generate all possible combinations of real or imagined objects, events, or situations) we obtain these possibilities:

1. None of the three hypotheses is correct.
2. The sight hypothesis is correct.
3. The smell hypothesis is correct.
4. The magnetic-field hypothesis is correct.
5. Both the sight and the smell hypotheses are correct.
6. Both the sight and the magnetic-field hypotheses are correct.
7. Both the smell and the magnetic-field hypotheses are correct.
8. All three hypotheses are correct.
9. Some other hypothesis is correct.
10. Some combination of other hypotheses is correct.

Having generated all possible combinations of likely hypotheses, the next task is to test them. Hasler tested the sight hypothesis first. To do so, he captured salmon that had just returned to two freshwater streams near Seattle, Washington – the Issaquah and East Fork. He then tagged the captured fish identifying which stream they had come from. Next, he randomly split the tagged Issaquah fish into two groups and blindfolded each fish in one group. He then repeated the procedure for the tagged East Fork fish. The blindfolded Issaquah and East Fork fish became the experimental group. Hasler then released the experimental fish along with some non-blindfolded fish (the control group) from both streams about three quarters of a mile below the junction where the streams join. Finally, the tagged fish were recaptured in traps about a mile above the junction as they swam back upstream. The following argument summarizes Hasler's reasoning:

> *If* . . . salmon find their home stream by sight (sight hypothesis),
> *and* . . . a group of non-blindfolded salmon and a group of blindfolded salmon from the Issaquah and East Fork streams are released below the fork where the two streams join (planned test),
> *then* . . . the non-blindfolded salmon should be recaptured in their home stream more frequently than the blindfolded salmon (prediction).

Logicians refer to this *If/and/then* reasoning pattern as deduction (e.g., Tidman & Kahane, 2003). Importantly, in this case the conclusion of the reasoning (the prediction or expectation) follows only when the planned test is controlled. In other words, to establish a link between the planned test's independent variable (i.e., the salmon's ability to see) and its dependent variable (i.e., where they are recaptured), all the other ways that the two groups of fish differ (all other possible independent variables) must be held constant. The reasoning sub-pattern, or sub-argument, that guides the construction of such "controlled" experiments is often referred to as the identification and control of variables (cf. Inhelder & Piaget, 1958).

Suppose having conducted this controlled experiment, we discover that the sighted salmon are better at returning home than the blindfolded salmon. Because this is the predicted result based on the sight hypothesis, the sight hypothesis would be supported, i.e.:

> *If* . . . the sight hypothesis is correct (sight hypothesis),
> *and* . . . the experiment is conducted as planned (planned test),
> *then* . . . the sighted salmon should be recaptured in their home stream more frequently than the blindfolded salmon (prediction).
> *And* . . . the sighted salmon were recaptured in their home stream more frequently than the blindfolded salmon (result).
> *Therefore* . . . the sight hypothesis is supported (conclusion).

However, as mentioned, one needs to be careful. Perhaps during the experiment, the blindfolded salmon were hindered in returning, not by lack of sight, but by their inability to swim with blindfolds. Or perhaps simply blindfolding the fish shocked them and disrupted their swimming ability.

Therefore, because one can never be certain that all such problems (i.e., all alternative explanations) have been eliminated all experimental results must be interpreted with caution.

On the other hand, suppose we find that the non-blindfolded and blindfolded salmon are equally successful at returning home. In this case, the sight hypothesis would not be supported. However, again one needs to be cautious. Overlooked independent variables might be operating. For example, perhaps the blindfolded salmon could see under their blindfolds. Or perhaps the blindfolds were too thin to block out all the light. Or perhaps the blindfolds were effective and the salmon do use sight when they can, but when they cannot, they use some other sense to navigate, such as smell. In short, the reasoning involved in experimentally testing hypotheses utilizes an *If/and/then/And/Therefore* pattern when results match predictions and an *If/and/then/But/Therefore* pattern when they do not. Because the point of both arguments is to test hypotheses via the deduction of predictions, the overall argument is referred to as hypothetico-deductive (HD) or sometimes hypothetico-predictive (cf. Cohen & Nagel, 1934; Popper, 1959, 1965; Platt, 1964; Chamberlain, 1965; Hempel, 1966; Medawar, 1969; Lawson, 2000; Lewis, 1988; Moore, 1993). The reasoning also involves identifying and attempting to control independent variables. However, because one can never be certain that all independent variables have been controlled, conclusions must remain somewhat tentative. Therefore, HD arguments and evidence can be convincing beyond a reasonable doubt; but they cannot be convincing beyond all possible doubt. In short, proof and/or disproof of any particular hypothesis are not possible.

As it turned out, when Hasler conducted the experiment, he found that the blindfolded salmon were as successful as the non-blindfolded salmon at finding their home streams. Therefore, the sight hypothesis was not supported. So Hasler moved on to test the smell hypothesis. To do so he again captured and tagged salmon from the two streams and randomly divided the Issaquah fish into two groups. He inserted cotton plugs coated with petroleum jelly in the noses of the experimental group fish to block their smelling ability and he left the noses of the control group unplugged. Hasler then randomly split the East Fork fish into two groups and plugged the noses of one group as he had done with the Issaquah fish. Finally, he released all the fish at the release point. As the fish returned upstream, they were recaptured in traps above the streams' junction. Of course, if the smell hypothesis is correct, then the smellers should end up in their home streams. Some non-smellers may get lucky and end up in their home stream as well. In fact, because there are only two streams, the use of probabilistic reasoning (i.e., the reasoning pattern used to identify and solve problems involving quantitative probabilistic relationships) suggests that half of the non-smellers will swim into their home stream by chance alone. So based on the smell hypothesis and probabilistic reasoning Hasler predicted that all of the smellers and half the non-smellers would be recaptured in their home streams.

During this experiment, 46 Issaquah control group fish were recaptured in the Issaquah and none were recaptured in the East Fork. Eight East Fork control group fish were recaptured in the Issaquah and 19 in the East Fork. In other words, $46 + 19 = 65$ of the 73 smellers returned to their home streams. Using proportional reasoning (i.e., the reasoning pattern used to identify and solve problems involving quantitative proportional relationships) we find that this ratio of 65/73 equals an 89% success rate. How does this success rate compare to the non-smellers? Among the experimental group, 39 of 51 Issaquah fish were recaptured in their home stream, as were 3 of 19 of the East Fork fish. So a total of $39 + 3 = 42$ of the 72 smellers ended up in their home streams. Again using proportional reasoning we find that this ratio of $42/72$ represents a 60% success rate.

The predicted percentages based upon the smell hypothesis were 100% for the smellers and 50% for non-smellers. So the observed percentages of 89% versus 60% are not exactly as predicted. However, suppose smell contributes nothing to stream-finding ability. If so, then the percentage for the smellers and the non-smellers should be the same. So the question we need to ask is this: is the 89% success rate of the smellers significantly higher than the 60% success rate of the non-smellers? Although a statistical analysis would be helpful, you might sense that, given the relatively large number of fish involved, and given the use of correlational reasoning, (i.e., the reasoning pattern used to identify and determine the extent to which two variables within a sample co-vary) the 89% appears to be substantially higher than 60%. In other words, the difference between the smellers' and the non-smellers' success rates is probably caused by the difference in smelling ability, rather than by chance. Therefore, we conclude that the smell hypothesis has been supported. Let's leave discussion of the magnetic-field hypothesis for another time.

Summary of Key Reasoning Patterns

The Hasler case study paints scientific inquiry and scientific reasoning as a process that seeks causes for puzzling observations. The process is a creative one that consists of identifiable components. First is the identification of the puzzling observation. Next is the use of analogical reasoning (abduction) to generate one or more hypotheses. Combinatorial reasoning may then be used to generate a list of all possible combinations of hypotheses. Typically, hypotheses are tested in order of most to least plausible by deducing one or more specific predictions based on the hypothesis under test (i.e., the “working” hypothesis) and its planned test. In general, the scientist’s goal is to find evidence in support of one or more of the hypotheses and evidence against the others. In other words, although each planned test must, in theory, render the working hypothesis “falsifiable” in the sense that contradictory evidence must be possible (Popper, 1959) scientists do not set out with the goal of rejecting their favored hypotheses (Woodward & Goodstein, 1996). In experimental contexts, the generation of planned tests requires another reasoning pattern referred to as the identification and control of variables. Finally, planned tests are conducted and data are collected and analyzed. Data analysis typically requires use of additional reasoning patterns such as probabilistic, proportional, and correlational reasoning.

Of course not all hypotheses are tested with experimental evidence. Circumstantial evidence can be used (e.g., why does this skull have pointed teeth? If the teeth are pointed because the animal is a predatory carnivore, and we look at its eye sockets, then they should be directed forward – to afford good depth perception needed to capture prey.). And correlational evidence can test hypotheses (e.g., why do some women with breast implants also have connective tissue disease? If having breast implants causes connective tissue disease, and we compare disease incidence among matched samples of women with and without implants, then the incidence of disease should be significantly higher in the implant sample than in the non-implant sample.).

Scientific reasoning can be further complicated the fact that much of science deals with the generation and test of theories that are more complex, more general, and more abstract than the hypotheses tested by Hasler. Theory testing requires a similar HD reasoning pattern but is often complicated by the fact that theories typically include the postulation of non-perceptible entities, thus require an additional argument, sometimes called a theoretical rationale (Lawson, 2003a), or warrant (Toulmin, 1958; Toulman, Rieke & Janik, 1984), to link the postulate under test with its deduced consequence. For example, to pit spontaneous generation theory, with its imagined vital force, against biogenesis theory in the 1700s, Lazaaro Spallanzani added seeds and water to several bottles and then heated their contents (De Kruif, 1953). He heated some for only a few minutes and boiled some for an hour. He then sealed their necks with a flame. For a control, he repeated the procedure with another set of bottles that he only corked. The reasoning plus the theoretical rationales, guiding Spallanzani’s experiment can be summarized like this:

> *If* . . . a vital force enters nonliving matter to bring it to life (spontaneous generation theory),
> *and* . . . the experiment is conducted as planned (planned test),
> *then* . . . after several days, microbes should be found in all the bottles (prediction). All of the bottles should contain microbes because the vital force should be able to enter them regardless of length of heating or method of sealing (theoretical rationale).

Alternatively,

> *if* . . . the vital force does not exist (biogenesis theory),
> *then* . . . microbes should be found in all the corked bottles and in the sealed bottles that were heated for only a few minutes, but not in the sealed bottles that were boiled for an hour (alternative prediction). This alternative prediction follows because, according to biogenesis theory, microbes can enter a bottle through or around a cork, but not through a sealed neck. Further, living microbes can survive a short period of heating but not an hour of boiling (theoretical rationale).

As you can see, the independent variable in Spallanzani's experiment was the presence or absence of corks, while the theoretical entity tested was the existence or non-existence of an imagined vital force. Thus, as this example shows, theory testing is seldom, if ever carried out by the direct manipulation of the postulated entity. Instead, theory testing involves indirect tests and requires the addition of one or more theoretical rationales to link manipulated experimental variables with theoretical postulates. The next section will consider the roles played by different brain regions during scientific reasoning and discovery.

THE NEUROLOGICAL BASIS OF REASONING

Let's take a brief look at the neurological basis of reasoning by considering how the brain processes visual input as this is the most well researched and understood area of neural functioning (discussion based in part on (Lawson, 2003b, Chapter 2)). We will then consider the nature of working memory and lastly turn to another case study of scientific reasoning to see how the identified brain regions may be involved.

Visual Recognition

Most people would guess that the brain processes visual input primarily in an inductive way – that is we look and we look again, and perhaps look still again, until we eventually induce an idea about what we are looking at. But this is not how the brain works. Instead, based on the initial look, the

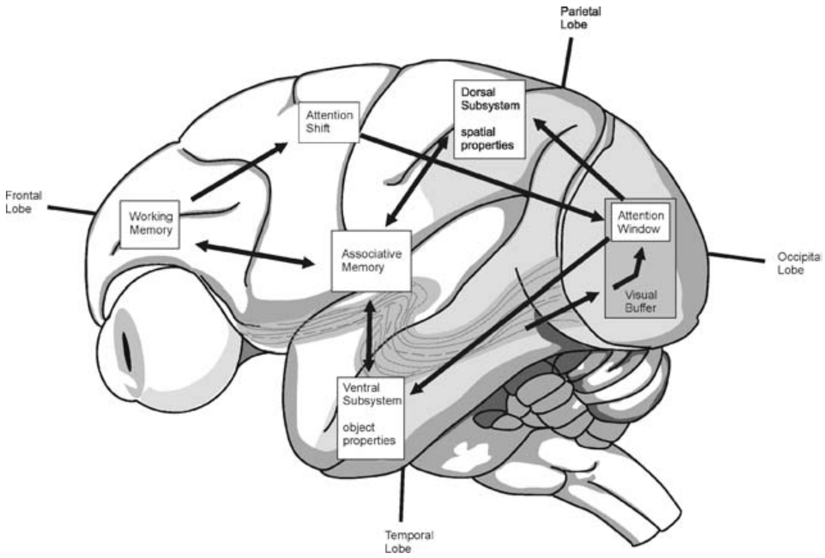


Figure 2. Kosslyn and Koenig's (1995) model of the visual system consists of six major subsystems. The order in which information passes from one subsystem to the next is shown. The subsystems generate and test hypotheses about what is seen in the visual field.

brain spontaneously and subconsciously generates a hypothesis of what might be out there and then uses subsequent looks to test its initial hypothesis. As summarized by Kosslyn and Koenig (1995), the ability to visually recognize objects requires participation of the six major brain regions (subsystems) shown in Figure 2. First, sensory input from the eyes produces a pattern of electrical activity in a region referred to as the visual buffer located in the occipital lobe at the back of the brain. This activity pattern produces a spatially organized "image" within the visual buffer. Next, a smaller region within the visual buffer, called the attention window, performs detailed processing. The activity pattern in the attention window is then simultaneously sent along two pathways on either side of the brain, one that runs down to the lower temporal lobe, and one that runs up to the parietal lobe. The lower temporal lobe, or ventral subsystem, analyses object properties, such as shape, color and texture, while the upper parietal lobe, or dorsal subsystem, analyses spatial properties, such as size and location.

Outputs from the ventral and dorsal subsystems come together in associative memory, which, as mentioned, is located primarily in the hippocampus, the limbic thalamus and the basal forebrain. The ventral and dorsal subsystem outputs are matched to patterns stored in associative memory. If a good match is obtained between inputs and stored conceptions, then the observer knows the object's name, categories to which it belongs, sounds it makes and so on. If not, the object remains unrecognized and additional sensory input must be obtained.

Importantly, the search for additional input is not random. Rather, stored patterns are used to make a second hypothesis about what is being observed, and this hypothesis leads to new observations and to further encoding. According to Kosslyn and Koenig (1995), "One actively seeks new information that will bear on the hypothesis. The first step in this process is to look up relevant information in associative memory." (p. 57) The information search involves activity in the prefrontal lobes in a region referred to as working memory. Activating working memory causes an attention shift of the eyes to a location where an informative component should be located. Once attention is shifted, the new visual input is processed in turn. The new input is then matched to shape and spatial patterns stored in the ventral and dorsal subsystems and kept active in working memory.

The key point with respect to reasoning is that visual recognition is a process in which the brain spontaneously generates and tests visual hypotheses in an HD or "top–down" manner. For example, suppose Joe, who is extremely myopic, is rooting around the bathroom and spots the end of an object that appears to be a shampoo tube. In other words, the nature of the object's end and its location prompt the spontaneous generation of a shampoo-tube hypothesis. In psychological terms, the visual input has been assimilated by Joe's shampoo-tube mental "structure" or mental "model" (cf. Grossberg, 1982; Johnson-Laird, 1983; Johnson-Laird, 2003; Piaget, 1985). Based on this initial hypothesis, as well as knowledge of shampoo tubes stored in associative memory (i.e., his shampoo-tube mental model), when Joe looks at the other end of the object, he expects to find a cap: *If* ... it really is a shampoo tube, *and* ... I look at the other end, *then* ... I should see a cap. Thus Joe shifts his gaze to the other end. *And* ... upon seeing the expected cap, he concludes that the object is in fact a shampoo tube. Importantly, other brain systems such as those involved in word recognition process information in a similar HD manner (see, for example, Kosslyn & Koenig, 1995, Chapter 6).

Everyday Hypothesis Testing

Simple "everyday" hypotheses appear to be tested in the same way. For example, consider the following question and response from a novice golfer:

Question:	Suppose length of a particular golf hole is listed at 156 yards. That length is measured from the beginning to what point on the green?
Novice golfer:	It's measured to the hole. No, that can't be right because they change the hole location each day. So I guess the length is measured to the center of the green.

Why do you suppose the novice golfer changed her mind? In other words, why did she reject her to-the-hole hypothesis so quickly after generating and stating it? Perhaps she subconsciously and instantaneously generated the following HD argument:

> *If* . . . the length is measured from the beginning to the hole location (to-the-hole descriptive hypothesis),
> *and* . . . they change the hole location each day (fact recalled from associative memory),
> *then* . . . the listed length would have to change each day (prediction).
> *But* . . . the listed length does not change each day (another fact recalled from associative memory).
> *Therefore* . . . the to-the-hole hypothesis is probably wrong and I'll need to generate another hypothesis (conclusion). Perhaps the length is measured to the center of the green!

Granted, we have no way of knowing for sure if she really did reason in this way. Nevertheless when subsequently shown this argument, the novice golfer thought that it made sense and that as far as she could tell, it was an accurate reconstruction of her reasoning.

Compare the previous golfer's response to that of a second novice golfer. When asked the same question, the second golfer simply guessed that the length is measured to the hole location and left it at that. When next asked if she knew that hole locations are changed each day she said no. So it appears that lacking this knowledge, she could not construct the HD argument that would have led her to reject her initial hypothesis and generate another one. Her failure to use HD reasoning is informative as it further clarifies the role of declarative knowledge in reasoning. Without knowing that hole locations change each day, the second golfer lacked key knowledge that could have been linked to the initial hypothesis and could then have led deductively to a contradiction. Hence, she did not make the deduction, did not discover the contradiction, and did not reject her initial hypothesis. We next turn to Galileo Galilei's discovery of Jupiter's moons in 1610 and attempt to identify the reasoning patterns involved as well as the roles played by the previously identified brain regions.

Galileo's Discovery of Jupiter's Moons

Lawson (2002) identified several cycles of HD reasoning that may have guided Galileo Galilei's discovery of Jupiter's moons in January of 1610. For example, when Galileo made his initial observation of three bright lights near Jupiter on the evening of January 7th, he immediately thought he was seeing three stars that were presumably embedded in the "fixed" celestial sphere behind Jupiter. However, his continued thinking led to doubt as revealed by the following remark:

> . . . and although I believed them to belong to the number of the fixed stars, yet they made me somewhat wonder, because they seemed to be arranged exactly in a straight line, parallel to the ecliptic, and to be brighter than the rest of the stars, equal to them in magnitude. ((Galilei, 1610), as translated and reprinted by Shapley, Rapport and Wright in 1954.)

Why would this observation lead Galileo to “somewhat wonder”? Of course we cannot know for certain, but he may have been reasoning along these lines (as suggested in Lawson, 2002): *If* the three bright lights are fixed stars, *and* their sizes, brightness and positions are compared to each other and to other nearby stars, *then* variations in size, brightness and position should be random, as is the case for other fixed stars. *But* “they seem to be arranged exactly in a straight line, parallel to the ecliptic, and to be brighter than the rest of the stars.” *Therefore* the fixed-stars hypothesis is not supported. Or as Galileo put it, “yet they made me somewhat wonder.”

Subsequently, Galileo rejected his fixed-stars hypothesis and presumably used other cycles of HD reasoning to also reject an astronomers-made-a-mistake hypothesis and then find support for the hypothesis that the bright lights were moons orbiting Jupiter, i.e.: *if* the three bright lights are orbiting moons, *and* I observe them over several nights, *then* some nights they should appear to the east of Jupiter and some nights they should appear to the west. Further, they should always appear along a straight line on either side of Jupiter. *And* this is precisely how they appeared. *Therefore* the moons hypothesis is supported.

The Role and Limits of Working Memory

Presumably Galileo’s cycles of HD reasoning took place in his working memory. Although research indicates that working memory is seated in the lateral prefrontal cortex, its location cannot be pinned down to a single prefrontal region. Rather its location appears to depend in part on the type of information being processed. With its many projections to other brain areas (including projections to those prefrontal neurons that store procedural rules such the controlling variable “rule”), working memory plays a crucial role in keeping representations active while coordinating mental activity.

Following Baddeley (1995), working memory, at least in adults, consists of three components – a visuo-spatial scratchpad, a central executive, and a phonological loop. In Baddeley’s theory, the visuo-spatial scratch pad activates representations of objects and their properties, while the phonological loop does the same for linguistic representations. Thus working memory becomes a temporary network to sustain information

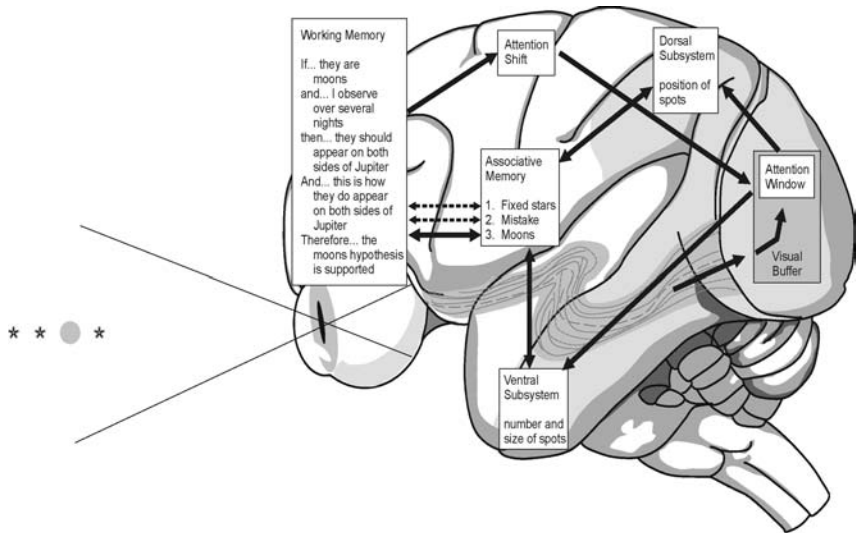


Figure 3. The contents of Galileo's working memory as he tests the moons hypothesis (from Lawson, 2002). Note how previously generated hypotheses stored in associative memory need to be inhibited (the dashed arrows) during the time the moons hypothesis, its predicted consequences, and the relevant evidence are active in working memory (the bold arrow).

while being processed. During reasoning, one must not only pay attention to task-relevant information but must also inhibit task-irrelevant information. Consequently, sustaining an argument in working memory involves allocating attention to and temporarily keeping track of information relevant to one's goals and actively inhibiting irrelevant information.

In terms of Galileo's reasoning and the Kosslyn/Koenig model, Figure 3 shows the possible contents of Galileo's working memory while testing the moons hypothesis. In order to draw a conclusion from the depicted argument, Galileo presumably must not only allocate attention to the hypothesis, to its planned test, to its predicted result, and to its observed result, he also must inhibit his previously generated fixed-stars and astronomers-made-a-mistake hypotheses. In short, he must keep a lot of relevant ideas in mind at the same time while suppressing potentially distracting ideas.

Importantly, it is now well known that working memory capacity is limited in terms of the independent "units" of data or thoughts that can be maintained at any one time. Miller's "magical" number 7, plus or minus 2, refers to the fact that it is almost universally true that people can recall only seven unrelated pieces of data (e.g., random letters or digits), if they do not resort to various memory tricks or aids (Miller, 1956; Pascual-Leone, 1970; Pascual-Leone & Ijaz, 1989). Clearly, however, we all form concepts that contain far more information than seven “units.” Thus, a mental process must occur in which previously unrelated units of input are grouped or “chunked” together to produce higher-order chunks (units of thought/concepts/rules). This implied process is known as chunking (Simon, 1974).

With respect to declarative knowledge, consider, for example, the ecosystem concept. An ecosystem is defined as a biological community plus its abiotic (non-living) environmental components. In turn, a biological community consists of producers, consumers, and decomposers; while the abiotic components consist of factors such precipitation, temperature, substrate type, and so on. Each of these subcomponents can in turn be further subdivided. Producers, for example, might include grasses, bushes, and pine trees. Thus, the ecosystem concept subsumes a far greater number of discrete units or chunks than seven. Thus, for those who have “constructed/chunked” the concept, it occupies but one unit or chunk in associative memory, and like procedural rules and peoples’ faces, it is likely “stored” in a single neuron. The result of chunking (i.e., of higher-order concept construction) is extremely important. Chunking reduces the load on mental capacity and simultaneously opens up additional capacity that can then be occupied by additional concepts. This in turn allows one to construct still more complex and inclusive concepts. Presumably, chunking also occurs during the construction of reasoning patterns (Wallis et al., 2001).

The key point in terms of complex cycles of HD reasoning is that sufficient chunking of concepts and/or reasoning sub-patterns needs to have occurred prior to one’s attempt to construct and maintain such complex arguments in working memory and use them to draw conclusions. If sufficient chunking has not occurred, then the reasoning (i.e., the HD argument) will “fall apart” and the conclusion will be “lost.” Importantly, this also means that such arguments and conclusions will also be lost on students who have not yet chunked the prerequisite concepts and/or reasoning sub-patterns such as those used by Hasler (e.g., control of variables, proportional reasoning, and correlational reasoning). This is why it is imperative that teachers know what their students bring with them to the classroom in terms of not only their levels of intellectual development but also their background of subject specific declarative knowledge. With these points in mind, we next turn to a brief look at the course of intellectual development with the intent of identifying the similarities and differences in reasoning patterns during childhood, adolescence and early adulthood.

THE COURSE OF INTELLECTUAL DEVELOPMENT

Children appear capable of a rudimentary form of HD reasoning virtually at birth. We can be fairly certain of this because the pattern can be found in non-humans. For example, Hauser (2000) conducted a revealing experiment with rhesus monkeys. First, a monkey was shown an eggplant – a favorite food item. In full view, the eggplant was then placed behind a screen. A second eggplant was then placed behind the screen. Then when the screen was lifted, the length of time the monkey looked at the two revealed eggplants was measured, which turned out to be about one second. Next the conditions were changed. In the initial changed condition, one eggplant was placed behind the screen followed by a second eggplant. Then without the monkey knowing it, the second eggplant was removed. Now when the screen was lifted, the monkey looked at the unexpected single remaining eggplant for about three to four seconds. The same increase in looking time occurred when a third eggplant was secretly added and then revealed. Thus, the monkey had a clear expectation of seeing two eggplants and when either one or three eggplants unexpectedly showed up, the monkey was puzzled as evidenced by the increase in looking time. In the first unexpected condition the monkey's "reasoning" can be summarized like this: *if* one eggplant is placed behind the screen, *and* another is added, *then* there should be two eggplants behind the screen. *But* there is only one eggplant. *Therefore* I am puzzled and need to look at the puzzling situation longer.

If we assume that this pattern of HD reasoning in humans is present at birth, then intellectual development involves a growing awareness (i.e., consciousness) of one's reasoning patterns and one's reflectivity as well increases in the contexts to which the patterns can be applied. Let's see how this might work in terms of Piaget's well-known concrete and formal operational stages of intellectual development (e.g., Inhelder & Piaget, 1958; Piaget & Inhelder, 1969) as well as a possible "post-formal" stage (Lawson, Clark, Cramer-Meldrum, Falconer, Kwon & Sequist, 2000a; Lawson, Drake, Johnson, Kwon & Scarpone, 2000b). Note that use of the Piagetian stage labels does not imply acceptance of his theory concerning their underlying operations (e.g., combinatorial system and INRC group).

The Concrete Operational Stage (7 Years to Early Adolescence)

Beginning at age seven, the prior acquisition of language to name objects, events and situations during the preoperational stage presumably allows the child to apply HD reasoning to a new level, the level of ordering and classifying, i.e., creating variables and higher-order classes/categories of

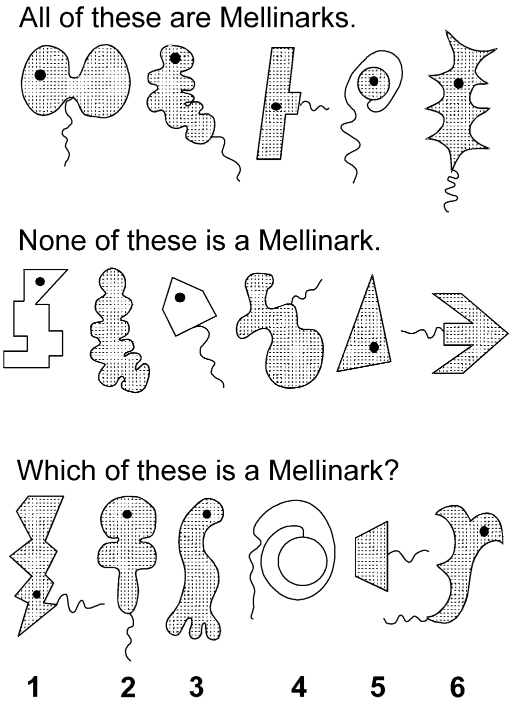


Figure 4. The Mellinarks (Elementary Science Study, 1974).

objects, events and situations. The observable and named objects such as tables and chairs of the preoperational stage become the classes/categories, such as furniture, of the concrete stage. For example, to test the hypothesis that concrete operational individuals are capable of constructing HD arguments to test descriptive hypotheses, a series of classification tasks, including the Mellinark Task (see Figure 4) were administered to children ranging in age from 6 to 14 years (Lawson, 1993). Brief one-on-one instruction was then used to teach them how to discover the relevant features using HD arguments, e.g.: *if* tiny spots make a creature a Mellinark (descriptive hypothesis), *and* I look at all of non-Mellinarks in row 2, *then* none of them should have tiny spots. *But* some do have tiny spots. *Therefore* tiny spots are not the key feature – or at least not the only key feature.

Interestingly, none of the six-year-olds could generate and/or comprehend this sort of argument, whereas half of the seven-year-olds could, as could virtually all of the eight to 14 year olds. Therefore, results supported the hypothesis that the concrete stage, which begins rather abruptly at seven years of age (most likely related to a growth spurt of the frontal lobes), involves the ability to use HD reasoning to serial order and to categorize the objects, events, and situations in the child's environment

– all mediated by language. That is, at the concrete stage, descriptive hypotheses are tested by comparing predictions with prior, preoperational, verbally-labelled constructions such as "spots," "tails," and "curvy sides."

The Formal Operational Stage (Early to Late Adolescence)

Following a comprehensive review of the psychological literature, Moshman (1998) concluded: "In fact, there is surprisingly strong support for Piaget's 1924 proposal that formal or hypothetico-deductive reasoning – deliberate deduction from propositions consciously recognized as hypothetical – plays an important role in the thinking of adolescents and adults but is rarely seen much before the age of 11 or 12" (p. 972). By hypothetical Moshman is referring to causal, as opposed to descriptive, hypotheses. For example, consider the question: what causes differences in the rates at which pendulums swing? To answer this causal question, one uses HD reasoning to generate and test alternative causal hypotheses (cf. Inhelder & Piaget, 1958, Chapter 4). For example: *if* changes in swing rates are caused by the amount of weight hanging on the end (causal weight hypothesis), *and* the weights are varied while holding other possible causes constant, *then* rate of pendulum swing should vary. *But* the rates do not vary. *Therefore* the weight hypothesis is not supported.

Clearly the reasoning pattern is the same as that used to test descriptive hypotheses during the prior concrete stage. Thus, the difference between formal reasoning and concrete reasoning is not the HD pattern. Again, the difference appears to be the context in which the pattern can be applied. Concrete reasoning is about testing descriptive hypotheses whereas formal reasoning is about testing causal hypotheses. The pendulum test involves an experiment in which the values of one possible cause are directly varied. Note also that the test involves use of the very important control of variables reasoning sub-pattern.

The Post-Formal or "Theoretical" Stage (Late Adolescence and Early Adulthood)

Consider once again Spallanzani's test of spontaneous generation and biogenesis theories. As you may recall, the HD argument summarizing his test went like this: *if* a vital force enters nonliving matter to bring it to life, *and* the experiment is conducted as planned, *then* after several days, microbes should be found in all the bottles. All of the bottles should contain microbes because the vital force should be able to enter them regardless of length of heating or method of sealing. Alternatively, *if* the vital force does not exist, *then* microbes should be found in all the corked bottles and in the sealed bottles that were heated for only a few minutes, but not in the sealed bottles that were boiled for an hour. This alternative prediction follows because, according to biogenesis theory, microbes can enter a bottle through or around a cork, but not through a sealed neck. Further, living microbes can survive a short period of heating but not an hour of boiling.

Although once again identical to prior reasoning in form, this argument differs from formal stage causal arguments in at least two important ways. As mentioned previously, here the proposed cause is unseen (i.e., theoretical) whereas at the formal stage, the proposed cause was observable. And unlike formal stage 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 Spallanzani's experiment, the independent variable is the presence or absence of corks, while the proposed cause is an unseen vital force or unseen microbes. Also as mentioned, because the proposed cause and the independent variable are not the same, a warrant, or theoretical rationale, is needed to link the two so that a reasonable test can be conducted. For these reasons, such reasoning is considered post-formal or theoretical, is more difficult than formal reasoning (e.g., Lawson et al., 2000a, 2000b), and is presumably not achieved until late adolescence after the final brain growth spurt at age 18 (Thatcher, Walker & Giudice, 1987; Thatcher, 1991), if at all.

Why is Intellectual Development Stage-Like?

Based on the previous arguments and evidence, we can understand why intellectual development is stage-like. In addition to probable maturational constraints, individuals construct something new during each stage that can be constructed only following the previous stage because the products of the previous stage are used in testing the possible constructions (i.e., the hypotheses) of the subsequent stage. For example, suppose we generate the theory that matter consists of tiny invisible and indivisible particles called atoms. Like John Dalton in the early 1800s, we can use post-formal reasoning to test this aspect of atomic theory, i.e.: *if* matter consists of invisible/indivisible particles that have specific weights and combine with one another in specific ways, *and* molecules are decomposed into their parts, *then* the ratios of weights of those parts should be in simple whole number ratios. *And* the ratios of weights of those parts are in simple whole number ratios. *Therefore* atomic theory is supported. Dalton's testing of atomic theory in this way required him to compare predicted and observed weight ratios of decomposed molecules. Comparing ratios involves proportional reasoning, a formal stage construction. Thus, Dalton's reasoning and eventual support for atomic theory could not have occurred without his prior construction of a proportional reasoning scheme.

Similarly, testing formal stage hypotheses requires use of prior concrete stage constructions. Consider Inhelder and Piaget's bending rods task (Inhelder & Piaget, 1958, Chapter 3). To test the causal hypothesis that variation in rod thickness causes variation in amount of rod bend (i.e., thinner rods bend more than thicker rods), one can reason like this: *if* differences in rod bending is caused by rod thickness, *and* equal weights are hung on two rods that vary only in thickness, *then* the thinner rod should bend more. *And* the thinner rod does bend more. *Therefore* the thickness hypothesis is supported.

Thus, to test the causal thickness hypothesis (one can directly observe/sense thickness differences), we must determine which of the two rods bends more and which bends less. In other words, we need to have already constructed a concrete stage "distance" variable, which we can label as "distance of bending." So to test a formal stage causal hypothesis, we use a prior stage construction (i.e., conservation of distance/length). Likewise, testing concrete stage descriptive hypotheses requires use of preoperational stage object-word constructs. And lastly, testing preoperational linguistic hypotheses requires use of sensory motor stage object constructs.

How Does Intellectual Development Occur?

How does procedural knowledge develop? We can answer this question in a general way by agreeing with Piaget that intellectual development occurs through self-regulation, i.e., by engaging in reasoning and by "internalizing" the products of that process and by internalizing (i.e., chunking) its procedures as well. According to Piaget (1976) a process he calls reflective abstraction provokes this internalization. Reflective abstraction progresses from the use of spontaneous actions to the use of explicit, verbally mediated rules to guide behavior. Reflective abstraction occurs when individuals are prompted by contradictory feedback (i.e., *If/and/then/But*) and the resulting state of mental "disequilibrium," to reflect first on their actions and later on arguments with others. Thus, the cause of reflective abstraction is contradiction by the physical environment and verbally by other people. The result of reflective abstraction is that the individual gains declarative knowledge and also becomes more aware of, more conscious of, more reflective, and more skilled in use of the procedures used in gaining that knowledge (i.e., declarative knowledge gets chunked and so does procedural knowledge).

This view of intellectual development helps clarify why "stage retardation" occurs, i.e., why some students fail to develop formal and post-formal reasoning patterns. Suppose, for example, years ago two isolated islands existed, each ruled by an all-powerful king. When questions arose, the islanders asked the king for answers – answers that were accepted as true. One day a foreign ship arrived at one of the islands. Over time, trading relationships were established between the island and several foreign countries. Importantly, not only did the ships bring new goods, the sailors also brought new ideas. The ideas spread throughout the island, some of which contradicted the "truths" previously handed down by the king. So the islanders began wondering which ideas were true, and more importantly, how they could tell. Eventually, an upheaval took place in which the king was overthrown and replaced by a government run by the people. Decades later, an anthropologist arrived on the island to study its culture. As part of her study, she administered a reasoning test to the island's adults. Soon after, she discovered the other island. She was the first "outsider" to discover the island, which was still controlled by an all-powerful king. She administered the reasoning test to the adults on this island as well. Which population of islanders do you think did better on the reasoning test? Clearly, the adults on the first island should be better. Piaget pointed out the reason as early as 1928 when he stated that the development of reasoning occurs as a consequence of "the shock of our thoughts coming into contact with others, which produces doubt and the desire to prove" (Piaget, 1962). 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. . . . 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. (p. 204)

In other words, the growing awareness of and ability to use internalized arguments to guide one's reasoning and decision making occurs as a consequence of attempting to engage in arguments of the same sort with others in which alternative hypotheses are put forward and accepted or rejected as the basis of evidence and reason as opposed to authority or emotion. If alternative ideas do not exist, then no external arguments ensue, and no internalization of patterns of argumentation results.

HOW CAN TEACHERS ENCOURAGE INTELLECTUAL DEVELOPMENT?

Given that several studies have found that many secondary school and college students have yet to develop formal and/or post-formal reasoning patterns and that their reasoning deficiencies lead to difficulties in problem solving, in understanding theoretical concepts, in rejecting misconceptions, and also to rejecting misconceptions about the nature of science and mathematics (as reviewed by Lawson, 2003b), more emphasis on teaching students to reason effectively is urged. Because effective reasoning lies at the very heart of scientific literacy, the key pedagogical question is this: How can we help more students develop formal and then post-formal reasoning patterns?

Teaching a Specific Scientific Procedure

To understanding of how instruction can provoke development of a specific formal reasoning sub-pattern, a specific procedure, consider controlled experimentation. Young children know when a previous test is "fair" or "not fair" when the variables are familiar; however they lack a general plan to set up "fair" tests ahead of time and in unfamiliar contexts (Wollman, 1977). For example, suppose two children run a race. The loser understands that the test was "unfair" if she was wearing heavy boots while the winner was wearing running shoes. Let's discuss in some detail how this intuition can be transformed into an internally mediated procedure in 9 and 13-year-old children who were initially unable to control variables. Following four half-hour individual training sessions, these same children were clearly able to demonstrate skill in controlling variables systematically and, in most cases, unhesitatingly. Further, as evidence of general skill in using this procedure, their skill transferred to new tasks, both manipulative and pencil-paper (Lawson & Wollman, 1976, 2003).

Session 1. Session one began by introducing each subject (S) to the intent and format of the instruction. Ss were told that materials would be used to teach them how to perform "fair tests." The materials were familiar: rubber balls, cardboard, foam rubber, a table. Ss were told that the first problem was to see which ball was the bounciest. To find out, they would instruct the experimenter what to do. Ss began by telling the experimenter to drop two balls to see which bounced higher. The experimenter would then drop two balls, but drop them from different heights. Ss would then respond by saying: "That isn't fair. Drop them from the same height." On the next trial the height would be equalized, however, one ball was dropped so that it hit the tabletop while the other hit the floor. This procedure was followed by continually intervening with new uncontrolled variables (spin one ball, push one ball, let one ball hit cardboard or foam rubber). Ss were then given a verbal rule – a test is "fair" if all the factors that might make a difference were the same for both balls. And tests in which these factors differed were called "unfair." The intent was to allow each S to generate a testing procedure, which was then contradicted. Presumably the contradictions forced Ss to reflect on the inadequacies of their self-generated procedures.

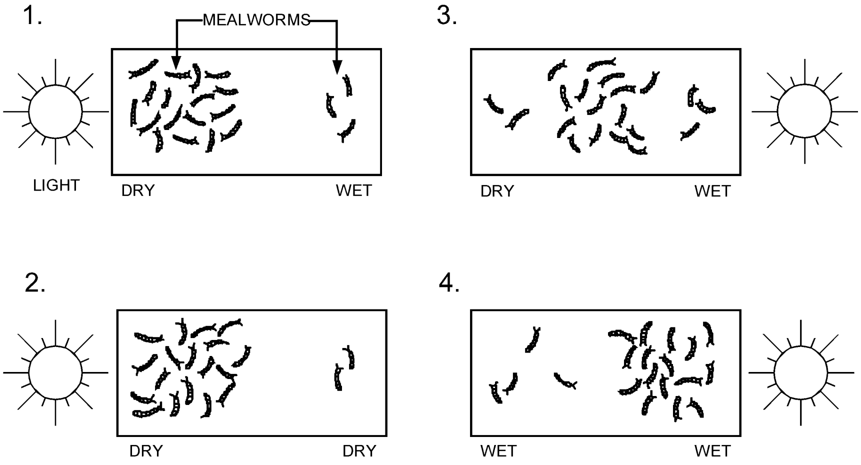
Session 2. Metal rods of varying size, shape, and material were placed on the table and each S was asked to classify them in as many ways as possible to insure that differences were noted. The rods were then placed into a stationary block of wood and factors that might affect the amount of bending were discussed. Ss were then asked to perform “fair tests” to find out if the identified factors did in fact affect the amount of bending (cf. Inhelder & Piaget, 1958). Whenever Ss performed a test, they were asked: Is it a fair test? Why is it fair? Questions were used to focus attention on the relevant variables, to recognize ambiguous experiments, and to understand the need for a procedure that keeps “all factors the same” except the one being tested. A number of examples and counter-examples were discussed at length.

Session 3. Ss were asked to experiment with an apparatus that consisted of a base that holds a post and an arm that attaches to the post. When pushed or propelled by a wound rubber band, the arm spins around like the rotor on a helicopter. Metal weights can be placed at various positions along the arm. Ss were briefly shown how the apparatus works and asked to discover factors that affect the time the arm spins before coming to rest. Following exploration, Ss were asked to perform “fair tests” to show that the factors mentioned do actually make a difference. Again, whenever a test was performed, Ss were asked questions that forced them to reflect on their procedures (e.g., Was it a fair test? Why was it fair?). The general intent of this session was similar to that of the second session as well as the fourth and final session. The strategies underlying the questions were identical in all sessions. The symbolic notation (the language used) remained invariant, while transformations in imagery were gained by first using familiar materials, and then by using unfamiliar materials. Ss were given a variety of tasks and were allowed to choose their own procedures. When mistakes were made, they were encouraged to reflect on their procedures and challenged to correct them.

Session 4. Two written problems (one of which appears in Figure 5) were used. The problems represented an additional step away from the physical and towards the verbal. Again probing questions relative to Ss’ understanding were asked. Thus, learning by doing was replaced solely by language.

Results. The four sessions clearly resulted in children who had internalized the meaning of the “fair test” rule. Importantly, they were capable of using the rule to design and conduct controlled experiments in novel contexts. Therefore, the results support the hypothesis that for intuitions

An experimenter wanted to test mealworms' response to differences in light and moisture. To do so he set up four boxes. He used lamps for light and placed watered pieces of paper in the boxes for moisture. In the middle of each box he placed 20 mealworms. One day later he counted the mealworms that had crawled to the ends of the boxes (see below).



Results show that mealworms respond (respond means move to or away from) to: (a) light but not moisture; (b) moisture but not light; (c) both light and moisture; (d) neither light nor moisture.

Figure 5. A written problem requiring the identification and control of variables.

to manifest themselves in the form of useful procedural rules, children need: (1) a variety of problems requiring a specific procedure for solution, (2) contradictions to their proposed procedures that force them to more closely attend to what they are doing or not doing, and (3) terms/phrases that remain invariant across transformations in materials – in this instance the key terms/phrases were "fair test" and "unfair test."

Hypothesis and Theory Testing in the Classroom

Although attempts at teaching formal and post-formal reasoning patterns in the classroom do not achieve the same degree of success as one-on-one sessions such as those just described, students do show marked improvements as a consequence of the right sort of classroom instruction (e.g., Cavallo, 1996; Germann, 1994; Harrison, Grayson & Treagust, 1999; Johnson & Lawson, 1998; Lawson, 1992, 1999; Lawson et al., 2000a, 2000b; Noh & Scharmann, 1997; Shayer & Adey, 1993; Shymansky, 1984; Shymansky, Kyle & Alport, 1983, 2003; Westbrook & Rogers, 1994; Wong, 1993; Zohar, Weinberger & Tamir, 1994). In general successful classroom instruction begins with explorations in which students encounter puzzling observations. For example, Lawson (1999) described a hypothesis-testing lesson that begins with a burning candle held upright in a pan of water using a small piece of clay. Shortly after a cylinder is inverted over the burning candle and placed in the water, the candle flame goes out and water rises in the cylinder. These puzzling observations raise two major causal questions: Why did the flame go out? And why did the water rise? The generally accepted answer to the first question is that the flame converted the oxygen in the cylinder to carbon dioxide such that too little oxygen remained to sustain combustion, thus the flame died. The generally accepted answer to the second question is that the flame transfers kinetic energy (motion) to the cylinder's gas molecules. The greater kinetic energy causes the gas to expand, which results in some escaping out the bottom. When the flame goes out, the remaining molecules transfer some of their kinetic energy to the cylinder walls and then to the surrounding air and water. This causes a loss of average velocity, fewer collisions, and less gas pressure. Water then rises into the cylinder until the total of the water and air pressure inside the cylinder is equal to the total of the atmospheric and water pressure outside the cylinder.

This lesson is a particularly good way to reinforce the idea that science is an alternative explanation generation and testing enterprise as the initial ideas students generate to explain why the water rises are experimentally contradicted, hence mental disequilibrium results along with the need for self-regulation. In other words, their ideas need to be replaced. A common student explanation centers on the idea that oxygen is "used up," thus a partial vacuum is created, which "sucks" water into the cylinder. Typically, students fail to realize that when oxygen "burns" it combines with carbon producing CO_2 gas of equal volume (hence no partial vacuum is created). Students also often fail to realize that a vacuum cannot "suck" anything. Rather the force causing the water to rise is a push from the relatively greater number of air molecules hitting the water surface outside the cylinder. Student experiments and discussions provide an opportunity to modify these misconceptions by introducing and testing a more satisfactory explanation of combustion and air pressure. An opportunity also exists to portray science as an intellectually stimulating and challenging way of using theories, in this case kinetic-molecular theory, to explain nature.

Although the present advocacy of such a hypothetico-deductive mode of instruction is not new (e.g., Heiss, Obourn & Hoffman, 1950; Van Deventer, 1958; Washton, 1967; Karplus & Thier, 1967; Lawson, Abraham & Renner, 1989), too few science curricular materials exist in which students openly generate and test hypotheses and theories. Of course, some lessons exist in which students are asked to test hypotheses. But in most of these cases, students are first told which hypothesis is correct and precisely how to test it. Such lessons probably do more harm than good as they teach little or nothing about how science is done, they encourage a reliance on authority, they lower motivation, and they encourage data fabrication (e.g., Lawson, Lewis & Birk, 1999). Therefore, what researchers and curriculum developers need to do is to design, evaluate, and disseminate more lessons in which students confront puzzling observations and are then challenged to generate and test alternative explanations. Importantly, the lessons must be accompanied by detailed teachers' guides that include not only lists of the hypotheses that have been generated, but also the hypothetico-deductive arguments and evidence used in their test.

CONCLUSION

In conclusion, the present paper paints successful human reasoning and scientific discovery in terms of cycles of hypothetico-deductive reasoning – reasoning in which working memory accesses and sustains hypotheses from associative memory to be tested and then actively seeks predictions and evidence that follow. In most instances, for most people, these reasoning cycles occur without conscious awareness. And unlike the streamlined arguments presented in this paper, the cycles often occur with many fits and starts. Nevertheless, successful reasoning follows the hypothetico-deductive pattern because the brain is "hard-wired" to process information in this way. But due to a variety of conditions, including lack of maturation of the frontal lobes, frontal lobe damage, or lack of relevant physical and social experience, reasoning abilities do not always develop to their full potential. Thus, failures may occur as evidenced by a lack of fruitful hypotheses to test, or more often, a premature acceptance of a pet hypothesis, often with little or no evidence in its favor. This leads to a failure to consider alternatives and potentially relevant evidence (e.g., Gabriel et al., 1994), or in terms of problem solving, a failure to consider and test alternative solution strategies – a condition often referred to as functional fixedness (e.g., Dominowski & Dallob, 1995).

The fact that not everyone develops formal and post-formal reasoning abilities carries important educational implications because higher-order reasoning is needed not only for decision making and for advanced problem-solving, but also to understand complex concepts and theories, to reject scientific misconceptions, and to understand the nature of science and mathematics – in other words for scientific literacy. The central implication is that if teachers hope to have their students become scientifically literate, they must teach in ways that allow students to develop the necessary reasoning abilities. In short, instruction must not only "fit" students' current developmental levels, but it must also provoke students to progress to higher levels. The evidence is clear that the best way to do this is to teach science in the way science is practiced. In other words, we should teach science as a process of critical inquiry where ideas are freely generated and rigorously tested.

ACKNOWLEDGEMENT

This material is based upon research partially supported by the National Science Foundation under grant No. DUE 9453610. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation. Thanks are due to James Shymansky and Larry Yore for their helpful comments on an earlier version of this manuscript.

REFERENCES

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

Baddeley, A. (1995). Working memory. In M.S. Gazzaniga (Ed.), *The cognitive neurosciences*. Cambridge, MA: MIT Press.

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

Cavallo, A.M.L. (1996). Meaningful learning, reasoning ability, and students' understanding and problem solving of topics in genetics. *Journal of Research in Science Teaching, 33*, 625–656.

Cohen, M.R. & Nagel, E. (1934). *An introduction to logic and scientific method*. London: Routledge and Kegan Paul.

Coll, R. & Treagust, D.F. (2002). Learners' use of analogy and alternative conceptions for chemical bonding: A cross age study. *Australian Science Teachers' Journal, 48*, 24–32.

Chamberlain, T.C. (1965). The method of multiple working hypotheses. *Science, 148*, 754–759. (First published in 1897.)

De Kruif, P. (1953). *Microbe hunters*. New York: Harcourt Brace & World. (First published in 1926.)

Dominowski, R.L. & Dallob, P. (1995). Insight and problem solving. In R.J. Sternberg & J.E. Davidson (Eds.), *The nature of insight*. Cambridge, MA: MIT Press.

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

Educational Policies Commission (1961). *The central purpose of American education*. Washington, DC: National Education Association.

Educational Policies Commission (1966). *Education and the spirit of science*. Washington, DC: National Education Association.

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

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

Gabriel, S.E., O'Fallon, W.M., Kurland, L.T., Beard, C.M., Woods, J.E. & Melton, I.I. (1994). Risk of connective-tissue disease and other disorders after breast implantation. *New England Journal of Medicine, 330*, 1697–1702.

Galilei, G. (1610). *The sidereal messenger*. In H. Shapley, S. Rapport & H. Wright (Eds.) (1954). *A treasury of science*. New York: Harper & Brothers.

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

Germann, P.J. (1994). Testing a model of science process skills acquisition: An interaction with parents' education, preferred language, gender, science attitude, cognitive development, academic ability, and biology knowledge. *Journal of Research in Science Teaching, 31*, 749–783.

Grossberg, S. (1982). *Studies of mind and brain*. Dordrecht, Holland: D. Reidel.

Hand, B.M., Prain, V. & Yore, L.D. (2001). Sequential writing tasks' influence on science learning. In P. Tynjala, L. Mason & K. Lonka (Eds.), *Writing as a learning tool: Integrating theory and practice*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hanson, N.R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge, England: Cambridge University Press.

Harrison, A.G., Grayson, D.J. & Treagust, D.F. (1999). Investigating a Grade 11 student's evolving conceptions of heat and temperature. *Journal of Research in Science Teaching, 36*, 55–87.

Hauser, M.D. (2000). What do animals think about numbers? *American Scientist, 88*, 144–151.

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

Heiss, E.D., Obourn, E.S. & Hoffman, C.W. (1950). *Modern science teaching*. New York: MacMillan.

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

Hoffman, R.R. (1980). Metaphor in science. In P.R. Honeck & R.R. Hoffman (Eds.), *The psycholinguistics of figurative language*. Hillsdale: 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. & Thagard, P.R. (1986). *Induction: Processes of inference, learning, and discovery*. Cambridge, MA: The MIT Press. (See especially Chapter 11.)

Inhelder, B. & 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: University of Chicago Press.

Johnson, M.A. & Lawson, A.E. (1998). What are the relative effects of reasoning ability and prior knowledge on biology achievement in expository and inquiry classes? *Journal of Research in Science Teaching, 35*, 89–103.

Johnson-Laird, P.N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

Johnson-Laird, P.N. (2003). Mental models and reasoning. In J.P. Leighton & R.J. Sternberg (Eds.), *The nature of reasoning*. New York: Cambridge University Press.

Karplus, R. & Thier, H.D. (1967). *A new look at elementary school science*. Chicago: Rand McNally.

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

Kosslyn, S.M. & Koenig, O. (1995). *Wet mind: The new cognitive neuroscience*. New York: The Free Press.

Lawson, A.E. (1992). The development of reasoning among college biology students. *Journal of College Science Teaching, 21*, 338–344.

Lawson, A.E. (1993). Deductive reasoning, brain maturation, and science concept acquisition: Are they linked? *Journal of Research in Science Teaching, 30*, 1029–1052.

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

Lawson, A.E. (2002). What does Galileo's discovery of Jupiter's moons tell us about the process of scientific discovery? *Science and Education, 11*, 1–14.

Lawson, A.E. (2003a). The nature and development of hypothetico-predictive argumentation with implications for science teaching. *International Journal of Science Education, 25*, 1387–1408.

Lawson, A.E. (2003b). *The neurological basis of learning, development and discovery: Implications for teaching science and mathematics*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Lawson, A.E., Abraham, M.R. & Renner, J.W. (1989). *A theory of instruction: Using the learning cycle to teach science concepts and thinking skills*. Cincinnati: National Association for Research in Science Teaching.

Lawson, A.E., Clark, B., Cramer-Meldrum, E., Falconer, K.A., Kwon, Y.J. & Sequist, J.M. (2000a). The development of reasoning skills 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. & Scarpone, C. (2000b). How good are students at testing alternative explanations of unseen entities? *The American Biology Teacher, 62*, 249–255.

Lawson, A.E., Lewis, C.M. Jr. & Birk, J.P. (1999). Why do students "cook" lab data? A case study of the tenacity of misconceptions. *The Journal of College Science Teaching, 29*, 191–198.

Lawson, A.E. & Wollman, W.T. (1976). Encouraging the transition from concrete to formal cognitive functioning – an experiment. *Journal of Research in Science Teaching, 13*, 413–430; also (2003). *Journal of Research in Science Teaching, 40*, S33–S50.

Lewis, R.W. (1988). Biology: A hypothetico-deductive science. *The American Biology Teacher, 50*, 362–367.

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

Miller, G.A. (1956). The magical number seven, plus or minus two. *Psychological Review, 63*, 81–97.

Moore, J.A. (1993). *Science as a way of knowing*. Cambridge, MA: Harvard University Press.

Moshman, D. (1998). Cognitive development beyond childhood. In D. Kuhn & R.S. Siegler (Eds.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language*, 5th edn. New York: Wiley.

Nauta, W.J.H. (1971). The problem of the frontal lobe: A reinterpretation. *Journal of Psychiatric Research, 8*, 167–187.

Noh, T. & Scharmann, L.C. (1997). Instructional influence of a molecular-level pictorial presentation of matter on students' conceptions and problem-solving ability. *Journal of Research in Science Teaching, 34*, 199–217.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 32*, 301–345.

Pascual-Leone, J. & Ijaz, H. (1989). Mental capacity testing as a form of intellectual development. In R.J. Samuda, S.L. Kong, J. Cummins, J. Lewis & J. Pascual-Leone (Eds.), *Assessment and placement of minority students* (pp. 143–171). Toronto: C.J. Hogrefe.

Paulesu, E., Frith, D.D. & Frackowiak, R.S.J. (1993). The neural correlates of the verbal component of working memory. *Nature, 362*, 342–345.

Piaget, J. (1962). *Judgment and reasoning in the child*. London: Routledge & Kegan Paul. (First published in 1928.)

Piaget, J. (1970). *Genetic epistemology*. New York: Norton.

Piaget, J. (1971). Problems of equilibration. In C.F. Nodine, J.M. Gallagher & R.D. Humphreys (Eds.), *Piaget and Inhelder on equilibration*. Philadelphia, PA: The Jean Piaget Society.

Piaget, J. (1976). Piaget's theory. In B. Inhelder & H.H. Chipman (Eds.), *Piaget and his school*. New York: Springer.

Piaget, J. (1985). *The equilibration of cognitive structures: The central problem of intellectual development*. Chicago and London: The University of Chicago Press.

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

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

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

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

Shayer, M. & Adey, P.S. (1993). Accelerating the development of formal thinking in middle and high school students IV: Three years after a two-year intervention. *Journal of Research in Science Teaching, 30*, 351–366.

Shymansky, J.A. (1984). BSCS programs: Just how effective were they? *The American Biology Teacher, 46*, 54–57.

Shymansky, J.A., Kyle, W.C. & Alport, J.M. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching, 20*, 387–404; also (2003). *Journal of Research in Science Teaching, 40*, S68–S87.

Simon, H.A. (1974). How big is a chunk? *Science, 183*, 482–488.

Squire, L.R. & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253*, 1380–1386.

Thatcher, R.W. (1991). Maturation of the human frontal lobes: Physiological basis of staging. *Developmental Neuropsychology, 7*, 397–419.

Thatcher, R.W., Walker, R.A. & Giudice, S. (1987). Human cerebral hemispheres develop at different rates and ages. *Science, 236*, 1110–1113.

Tidman, P. & Kahane, H. (2003). *Logic and philosophy*, 9th edn. Belmont, CA: Wadsworth/Thomson.

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

Toulman, S.E., Rieke, R. & Janik, A. (1984). *An introduction to reasoning*, 2nd edn. New York: Macmillan.

Van Deventer, W.C. (1958). A simplified approach to the problem of scientific methodology. *School Science and Mathematics, 58*, 99.

Wallis, J.D., Anderson, K.C. & Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411*, 953–956.

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

Washton, N.S. (1967). *Teaching science creatively in the secondary schools*. Philadelphia: Saunders.

Westbrook, S.L. & Rogers, L.N. (1994). Examining the development of scientific reasoning in ninth-grade physical science students. *Journal of Research in Science Teaching, 31*, 65–76.

Wollman, W. (1977). Controlling variables: Assessing levels of understanding. *Science Education, 61*, 371–383.

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.

Woodward, J. & Goodstein, D. (1996). Conduct, misconduct and the structure of science. *American Scientist, 84*, 479–490.

Zohar, A., Weinberger, Y. & Tamir, P. (1994). The effect of biology critical thinking project on the development of critical thinking. *Journal of Research in Science Teaching, 32*, 183–196.

School of Life Sciences,
Arizona State University,
Tempe, AZ 85287-4501,
U.S.A.
E-mail: anton.lawson@asu.edu