Teaching & Learning the Scientific Method

Guy R. McPherson

SCIENTISTS and science educators express concern about science literacy frequently. Better understanding of the scientific method is a common component of pleas for increased literacy. However, I believe that lack of understanding of the scientific method is more rampant than is commonly believed, at least in part because many scientists and science educators do not understand all the components of the scientific method. Specifically, misuse of the term "hypothesis" obfuscates genuine understanding of the scientific method. We routinely use the term "hypothesis" when we mean "prediction." This unacceptable substitution dilutes the power of the scientific method to the extent that invoking the "scientific method" has become largely meaningless.

One danger in discussing the scientific method is the implication that there is a single list of steps, or recipe, that generates reliable knowledge. Of course, no single series of steps could contain all the strategies that can be employed in the quest to understand the universe, and the "scientific method" reflects a classical philosophical perspective rooted in falsifiability (Popper 1968). This classical perspective has been substantially expanded by contemporary philosophers of science, notably in biology and ecology (Ruse 1979, 1988; Sober 1984; Thagard 1992; Pickett et al. 1994). Nonetheless, the widespread use and teaching of the "scientific method" suggests that this particular approach helps us organize our thoughts about the scientific process. If we are committed to using and teaching this set of steps—and apparently we are—there should be some agreement about how the steps are taught. The objective of this paper is to encourage consensus in teaching science. I begin by defining a few terms, then provide evidence for confusion about the scientific method, and conclude by illustrating some of the consequences of misuse of the term.

Observation and consequent description of a natural phenomenon generally initiate scientific inquiry. Observations are compared and patterns are sought, frequently with the assistance of statistical tools. For example, the observation that woody plants establish in greater numbers on a site underlain by loamy soil than on an adjacent site underlain by clayey soils may prompt a researcher to determine if this observation indicates a general pattern: s/he samples several sites underlain by each substrate, then constructs and tests a *statistical* hypothesis to determine the generality of the original observation. This process may be—and frequently is—repeated in several organisms, systems or regions by different investigators. Basic observations and descriptions contribute to the discovery and documentation of patterns, and each of these steps is fundamental to increasing our understanding of nature. However, the hypothetico-deductive method has not been employed to this point; rather, we have merely made observations. As such, we have neither generated nor tested *scientific* hypotheses, which—from the Popperian perspective—are candidate explanations for observed patterns (Medawar 1984: Matter & Mannan 1989). That is, a hypothesis is a potential reason for the pattern. Demonstration of a pattern often generates the question: "What process causes that pattern?" Providing a definitive answer to this question requires formulation and subsequent testing of potential explanations for the observed patterns; that is, it involves hypothesis testing. Cogent reviews of the hypothetico-deductive method are presented by Popper (1968) and Medawar (1984).

Guy R. McPherson is Professor in the School of Renewable Natural Resources at the University of Arizona, Tucson, AZ 85721.

Statisticians have been aware of the difference between *statistical* hypotheses and *scientific* hypotheses for many years: the former are used to identify or elucidate patterns, whereas the latter are used to identify mechanism(s) underlying pattern(s). Edwards (1972, p.180) echoed earlier statisticians in issuing a plea to understand and appreciate the difference: "What used to be called judgement is now called prejudice, and what used to be called prejudice is now called a null hypothesis." He calls such inappropriate use of the null hypothesis "dangerous nonsense (dressed up as 'the scientific method') [which] will cause much trouble before it is widely appreciated as such."

Distinguishing Hypotheses from Predictions

A prediction is a statement that is likely to be factual. Most predictions can be evaluated via observation, although the instruments of observation are variable and occasionally quite sophisticated. Consider, for example, the prediction "there are no living organisms on Mars." Evaluating this prediction requires sophisticated technology. It does not, however, require use of the scientific method, complete with hypothesis testing and formulation. Powers of observation, expanded by technology, must be used to determine whether there is life on Mars (or a particular species of fish in a stream, or a plant in a meadow). If this process qualifies as use of the scientific method, then the scientific method is not unique to science, but rather is used for everyday activities such as mowing the lawn (the grass is too long in some spots, so I must have "missed" those spots), shopping for groceries (I only need milk, so I will look in the dairy section rather than searching the entire store), and commuting to the workplace (length of route and traffic patterns dictate my path). In other words, science has little to offer beyond everyday activities if observation is the only means by which we acquire reliable knowledge.

Teachers can use relatively simple examples to illustrate and explain the difference between predictions and hypotheses. If field trips to two or more locations reveal the presence of a specific species (e.g. little bluestem, *Schizachyrium scoparium*), most students would predict that the species would be found at a nearby site with the same climate and similar elevation. Similarly, most students would predict a relationship between specific weather patterns and occurrence of wildfires, or between precipitation patterns and abundance of flowers. However, most students are unable to distinguish these predictions from hypotheses: that is, they formulate statements such as: "I hypothesize that little bluestem will be found on the next field trip" or "I hypothesize that there is a relationship between weather patterns and fire occurrence." In each case, "predict" is a more appropriate term than "hypothesize."

A follow-up field trip may reveal that little bluestem is *not* present in a location where students predicted it would occur, based on information about climate and elevation. The absence of little bluestem from a particular site where it was expected to occur naturally leads curious students to wonder why it is absent, and to formulate hypotheses (i.e. candidate explanations). Similarly, a strong relationship between weather patterns and other natural phenomena (e.g. wildfires, flowering) spurs explanations about the relationship. Testing a hypothesis requires the use of deductive logic to develop expectations in light of the proposed explanation. Students may generate hypotheses regarding little bluestem that deal with land use, soil fertility, presence of other species, parent material, or any number of other factors. They may generate mechanistic explanations about the relationship between climatic events and wildfires (e.g. precipitation during the growing season enhances growth of herbaceous fuels, which removes a primary constraint on fire occurrence and spread). The steps of formulation and testing hypotheses lead naturally to experimentation as a means of accessing cause and effect. In the case of little bluestem, students may propose experiments that manipulate land use, soil fertility, abundance of other species, or parent material (via reciprocal transfer of soils). Of course, experimentation is not ethically or logistically possible in all situations, which contributes to relatively weak inference (*sensu* Platt 1964). In contrast to hypothesis testing, experimentation generally is not required to assess predictions: predictions require observation, potentially aided by technology, but they do not require use of the "scientific method."

Hypotheses must be stated in a manner that makes them amenable to testing and falsification (Popper 1981). Similarly, most predictions can be stated in a testable and falsifiable manner (e.g. "there are no living organisms on Mars" or "little bluestem will not be found at the next location we visit"). Clearly, testing and falsification are insufficient criteria to differentiate between predictions and hypotheses.

Evidence of Confusion

I have had numerous discussions with my colleagues at a major research institution about this issue. Most of these discussions begin with a question I pose to Ph.D. students at their oral comprehensive examination: "What is a hypothesis?" When I first began asking this question more than 10 years ago, nearly all students invoked Popper's falsificationist view (e.g. Popper 1981), then proceeded to confuse hypothesis with prediction. Within the last five years, most students have been answering the question correctly, presumably because they have taken my classes, participated in my seminars, or talked to other students who helped them prepare an appropriate response. Although most students have learned to answer the question in a satisfactory manner, they still lack genuine understanding. Throughout this 10-year period, the question has invoked discussion from other faculty members involved in the examination, many of whom are as confused as the students being tested. A common argument is that my use of the term "hypothesis" is "too restrictive," an argument that would be moderately compelling if it were based on the writings of contemporary philosophers of science. Instead, it clearly has been based on a misunderstanding of the hypothetico-deductive method (i.e. "scientific method").

Another personal anecdote reinforces the view that scientists (including those outside my academic institution) may confuse "hypothesis" and "prediction." In a recent manuscript, my co-author and I substituted the inappropriate term "hypothesis" for the appropriate term "prediction" throughout the manuscript (a total of seven times). The paper was submitted to a major international journal in 1998 and was subjected to a high level of scrutiny by three reviewers (presumably scientists who are reasonably well known), the associate editor, and the editor-in-chief (the latter two individuals are internationally renowned scientists). None of the five individuals in the review process commented on the inappropriate use of "hypothesis," and the manuscript was published in the journal in 1999 (corrections were noted by the co-authors on the page proofs, and incorporated in the published manuscript).

Evidence far more compelling than these anecdotes can be found in virtually every issue of every journal. Consider, for example, a recent review of the scientific method in *The American Biology Teacher*, in which Sterner (1998) employed the statistical definition of "hypothesis" (including "null" and "alternative" hypotheses). Hypothesis, as used by statisticians, clearly refers to elucidation of a consistent pattern, not determination of causality. Hence, a statistical hypothesis and a scientific hypothesis are not equivalent, as Sterner implies. The scientific method is neither distinguishable from everyday activities nor particularly powerful when used in this manner because the mechanism explaining the pattern remains unknown, regardless of the tool used to observe the pattern. However, the scientific method, complete with experimentation, can be used to ascertain mechanism (e.g. the reason *why* two populations differ in some respect).

Misuse of the term "hypothesis" and the resulting misunderstanding of the scientific method are not restricted to *The American Biology Teacher*. I suggest that such misuse permeates the scientific literature and scientific searches for pattern. Readers who are interested in documenting this phenomenon need only look in the nearest journal: I recommend starting with *Science* or *Nature*.

Consequences

Formulating and testing scientific hypotheses are key components of the scientific method. Scientists (even adolescent ones) want to incorporate these components into their work. Explicit hypotheses offer clarity to presentations and papers, to the point that I suspect many scientists believe their science is second-class (or worse, not science at all) if they are not formulating and testing hypotheses. Further, patterns are easy to describe and assess relative to the difficult task of devising tests to differentiate between alternative candidate explanations. Consequently, there is great temptation to claim that hypotheses are being formulated and tested even when they are not (i.e. to expand the definition of hypothesis to include virtually any statement). One result is that virtually all contemporary biological research is said to test hypotheses. Of course, most research describes patterns rather than testing mechanisms underlying the patterns and therefore does not involve hypothesis testing. That is, most research does not employ the hypothetico-deductive method and is therefore not mechanistic in nature. Nonetheless, hypotheses (i.e. candidate explanations for observed patterns) are formulated in most published papers: they appear in the discussion, awaiting development of tests and execution of experiments for some later time.

One obvious consequence of the misuse of terminology is positive feedback among scientists: we are so committed to testing hypotheses (after all, that's what scientists do), that we have maligned the term so that we can all test "hypotheses." Perhaps a more important concern is that the general populace, whom we are trying so hard to educate, cannot determine what characteristics differentiate science from any other activity. After all, if a statement such as "there are two brands of milk in a neighborhood grocery store" is a hypothesis, then we certainly do not need to invoke a scientific activity (such as the "scientific method") to test hypotheses. In fact, this prediction can be evaluated by virtually anyone, which implies that everyone is practicing science all the time.

Learning Exercise

After differentiating between hypotheses and predictions in a lecture early in the semester, I select five to 10 articles from the primary journal literature and assign them as required reading for a near-future class discussion. I select recently published papers directly related to a topic that we have discussed or are about to discuss in class (e.g. ecological interactions, succession). Further, I select papers that represent the full range of approaches to the "scientific method," from those that document patterns to those that propose and experimentally test mechanistic hypotheses. Nearly all such papers use language suggesting that hypotheses are being tested.

When the papers are assigned, I ask students to read them with the following questions in mind:

1. Did the paper propose hypotheses or predictions?
2. Were the hypotheses or predictions clearly stated?
3. Were the hypotheses or predictions testable?
4. Were the hypotheses or predictions tested?
5. On a scale of one to five, how comfortable are you with the paper's conclusions?

I ask students to bring answers to these questions on a sheet of paper for each paper I asked them to read. I treat these responses anonymously and do not grade them. Obviously, they could be graded and used for evaluation.

Before we begin an in-class discussion, I tally the results, which nearly always suggest that students readily accept predictions as hypotheses; in other words, my lecture—which was designed to allow students to differentiate between the two—failed to produce the desired effect. This is not particularly surprising, given the limitations of lecturing as a teaching strategy. However, students have now tried to apply knowledge, and we use this attempt as a basis for subsequent discussion. Students usually are very engaged in these discussions, presumably in part because they (nearly all) failed to grasp the fundamental concept and they want to know why. This motivation, triggered by discomfort and satisfied by thorough discussion, engenders genuine learning: follow-up exercises and exams indicate that nearly all students can distinguish between hypotheses and predictions in journal articles after our in-class discussion.

I am not arguing that all scientists, or even all sciences, must regularly use the "scientific method" as it is usually described. Identification and elucidation of patterns are necessary steps in the quest to understand the natural world, and most scientists rarely have the opportunity to employ the hypothetico-deductive method. However, we should not confuse identification and elucidation of patterns (including *statistical* hypothesis testing) with *scientific* hypothesis testing. Pattern assessment is part of our daily lives and is also a fundamental part of science. In contrast, application of the scientific method, complete with hypotheses that can be formulated and tested, is one of relatively few activities unique to science. This method can be a powerful and valuable tool for determining processes underlying patterns, and understanding the scientific method creates a solid foundation for science literacy. If we insist on teaching the scientific method as a recipe, we should agree what it is.

Acknowledgments

Erika Geiger, Cindy Salo, Cody Wienk, Bill Matter and Bob Steidl graciously provided thoughtful reviews of an earlier draft of this manuscript.

References

Edwards, A.W.F. (1972). *Likelihood.* Cambridge, MA: Cambridge University Press.

Matter, W.J. & Mannan, R.W. (1989). More on gaining reliable knowledge: A comment. *Journal of Wildlife Management, 53*, 1172–1176.

Medawar, P. (1984). *Pluto's Republic.* New York: Oxford University Press.

Pickett, S.T.A., Kolasa, J. & Jones, C.G. (1994). *Ecological Understanding: The Nature of Theory and the Theory of Nature.* San Diego: Academic Press.

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

Popper, K.R. (1968). *The Logic of Scientific Discovery.* New York: Harper and Row.

Popper, K.R. (1981). Science, pseudo-science, and falsifiability. In R.D. Tweney, M.E. Doherty & C.R. Mynatt (Eds.), *On Scientific Thinking* (pp. 92–99). New York: Columbia University Press.

Ruse, M. (1979). Falsifiability, consilience, and synthesis. *Systematic Zoology, 28*, 530–536.

Ruse, M. (1988). *Philosophy of Biology Today.* Albany, NY: State University of New York Press.

Sober, E. (1984). *The Nature of Selection: Evolutionary Theory in Philosophical Focus.* Cambridge, MA: MIT Press.

Thagard, P. (1992). *Conceptual Revolutions.* Princeton, NJ: Princeton University Press.

Sterner, R.T. (1998). The scientific method: An instructor's flow chart. *The American Biology Teacher, 60*, 374–378.