Behavior and Philosophy, *29*, 1-20 (2001). © 2001 Cambridge Center for Behavioral Studies

THE WEAKNESSES OF STRONG INFERENCE

William O'Donohue
Jeffrey A. Buchanan
University of Nevada, Reno

ABSTRACT: Platt's 1964 paper "Strong Inference" (SI), published in *Science*, has had considerable influence upon the conception of the nature of the scientific method held in both the social and physical sciences. Platt suggests that a four-step method of specifying hypotheses and conducting critical experiments that systematically eliminate alternatives has been responsible for progress in the history of the successful sciences and, if adopted, will allow the less successful sciences to make greater progress. This paper criticizes SI on a number of grounds including: 1) no demonstration of the central historical claim, that is, that the more successful sciences actually have utilized this method more frequently than the less successful sciences; 2) poor historiography, in that more faithful explications of some of the historical case studies Platt cites as canonical examples of SI fail to support that SI was actually used in any of these cases; 3) failure to identify other cases of important scientific progress that did not use SI and instead used distinct scientific methods; 4) neglect of the importance and implications of background knowledge used in problem formulation; 5) the impossibility of enumerating a complete set of alternative hypotheses; 6) no acknowledgement of the Quine-Duhem thesis indicating that "critical" experiments are actually logically inconclusive; 7) the assumption that there is one scientific method; and 8) significant ambiguity regarding exactly how each of the four steps of SI is to be faithfully implemented. Some recommendations regarding a normative scientific method are given.
Key words: strong inference, hypothesis testing.

In 1964, John R. Platt's "Strong Inference" (SI) was published in *Science*.[1] This article described a method of scientific inquiry that was purported to be the method of successful experimental sciences (SI does not apply to nonexperimental sciences such as astronomy) and the means by which less successful sciences could progress more rapidly. SI has had a significant impact on the social and natural sciences. A search of the *Social Science Citation Index* since 1970 revealed that Platt's article has been cited 360 times in the social science literature and an additional 322 citations were found in the *Science Citation Index*. A few quotes from some of these articles may help illustrate the influence SI has had. For

AUTHORS' NOTE: Please address all correspondence to William O'Donohue, Ph.D., Department of Psychology/MS298, University of Nevada, Reno, Reno, NV 89557 (Email: wto@unr.edu) or to Jeffrey A. Buchanan, MA, Department of Psychology/MS298, University of Nevada, Reno, Reno, NV 89557 (Email: jeffreyb@scs.unr.edu).

[1] All references to Platt in this essay are to his Strong Inference (1964) article, hereafter cited in the text by page number.

example, Zimmerman and Zimmerman (1976) state that, "we have long accepted the inductive approach as good science (Platt, 1964), knowing that many experimental results can't be unambiguously accounted for until all possible interpretations are considered" (p. 393-394). Furthermore, " . . . when this occurs, multiple hypotheses that could account for the observation are then listed. The most acceptable hypothesis emerges only after crucial experiments have been conducted that eliminate all alternative hypotheses" (p. 394).

Roberts and Glick (1981), in explaining weaknesses in task design research, state one weakness as, "Little attempt has been made to provide strong inference research (Platt, 1964) in which various theories of task design are compared . . . " (p. 210). Massaro and Cowan (1993) state that for overcoming constraints in psychological inquiry, "First, the research strategies of falsification (Popper, 1959) and strong inference (Chamberlin, 1965; Platt, 1964) should be used" (p. 388). Folger (1989) asserts that, "The practice of good science clearly has a place for pitting theories against one another in what are regarded as crucial experiments. Anyone who remains unconvinced should read Platt's (1964) account . . . " (p. 159). Smith and Sechrest (1991) have a section devoted to "Using strong inference and testing alternative models" when discussing ways in which aptitude by treatment interaction research could be improved (p. 241). Mackenzie (1984) discusses three hypotheses that could be used to explain IQ differences in Black and White individuals. He states that strong inference methodology is the strongest method for gathering evidence about the relative influence of environmental versus genetics hypotheses (p. 1215-1216). Dixon (1998) proposes a methodology that " . . . can be considered an application of the method of strong inference" for progressing research in the area of developmental ordering (p. 144). Borkovec (1997), in advocating for a basic science approach to psychotherapy research, states that, "The optimal context for such scientifically focused research comes from strong inference . . . " (p. 146).

In addition to frequent citations in the "hard" science and social science literature, the SI method is endorsed in several psychology textbooks. These textbooks include a range of subdisciplines in psychology such as abnormal psychology (Kendall & Hammen, 1995; Mahoney, 1980; Rathus & Nevid, 1991), developmental psychology (Hertzog, 1994), and research methods (Bordens & Abbott, 1991; Ghosh, 1984).

It has also been our admittedly informal observation that this paper is one of the most frequently required readings in many graduate level methodology courses in psychology. Therefore, it appears that overall, SI has had a strong impact on the social sciences as a whole, including psychology.

Despite its apparent popularity and influence, we will argue that the methodological prescriptions and historical claims contained in SI are problematic. We will first outline Platt's main arguments and then examine the major problems associated with SI. We will suggest that, although Platt apparently presents a relatively simple, systematic, and efficient method of scientific inquiry, closer examination reveals that these putative characteristics are illusory. The appeal of these positive characteristics may be one reason that SI is cited frequently and praised as the regulative methodology, but at the same time is actually infrequently employed in practice in the social sciences, particularly in psychology. A recent search (through PsychInfo) of the psychological literature using the key words "strong inference" revealed that less than 10 empirical studies actually reported utilizing SI.

A Summary of Platt's Argument

SI involves four steps that Platt claims can be applied to any problem in science. These steps are:

> 1) Devising alternative hypotheses,
> 2) Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses,
> 3) Carrying out the experiment so as to get "clean" results,
> 4) Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain. (p. 347)

Platt uses the metaphor of climbing a tree: at each fork in a branch there are alternative hypotheses diverging in different directions. The results of critical experiments help determine which branch is taken by excluding one possible fork. This testing-elimination process continues until all alternatives but one are ruled out. It is in this sense that Platt views SI to be an inductive method of inquiry because results from particular cases (i.e., specific critical experiments) lead to general conclusions.

Historical Influences on SI

The method of SI as Platt describes it appears to have three major historical influences. The first is Francis Bacon. Platt sees Bacon's influence in that he states, "In its separate elements, SI is just the simple and old-fashioned method of inductive inference that goes back to Francis Bacon" (p. 347). Bacon argued that an inductive method of forming rival hypotheses and conducting critical experiments that exclude one hypothesis in favor of the other was a "surer method" of "finding out nature" (p. 349). In addition, the metaphor of the forking tree that Platt uses is usually credited to Bacon. Platt states (p. 349):

> Of the many inductive procedures he [Bacon] suggested, the most important, I think, was the conditional inductive tree, which proceeded from alternative hypotheses (possible "causes" as he calls them), through crucial experiments ("Instances of the Fingerpost"), to exclusion of some alternative and adoption of what is left ("establishing axioms").

The second and more contemporary influence on Platt is the philosopher of science Karl Popper. Popper emphasized that the scientist can never establish the truth of a theory by "empirical reasons" (Popper, 1981). That is, a theory cannot be "proven" through repeated empirical confirmations because no matter how many confirmations the scientist gathers, there are an infinite number of untried tests, any of which, in principle, could reveal results that are inconsistent with the theory. Einstein's eventual refutation of the often "confirmed" Newtonian theory illustrated for Popper how a very well confirmed theory can eventually be shown to be false. Another classic example showing the pragmatic difficulties involves the statement "All swans are white." This statement can never be "proven" to be true by any finite number of observations of white swans because it is not possible to observe every swan that has existed, currently exists, or that will exist in the future. However, Popper (1981) claims that the *falsity* of a theory can sometimes be established, assuming that the empirical states of affairs that the theory rules out can actually be revealed in an empirical test. Platt adopts this falsifiability criterion for evaluating scientific theories (see p. 350). For example, Platt states that in order to determine if a hypothesis is falsifiable, one should ask its proponent, "But sir, what experiment could disprove your hypothesis?" (p. 352).

The final historical influence on Platt is that of T. C. Chamberlin, with whom Platt credits the "method of multiple hypotheses." In this method, several alternative explanatory hypotheses are proposed, rather than one. Chamberlin and Platt argue that when a scientist proposes a single hypothesis, he can become psychologically attached to it. This allegiance can cause the scientist to construe any experimental findings as supporting the hypothesis. This cognitive tendency is related to "confirmation bias," in which people (including scientists) tend to interpret data in ways that support previously held beliefs or theories, even if the data appears to be clearly contradictory (see Nisbett & Ross, 1980 for a review of this literature). Chamberlin and Platt suggest that by devising several alternative hypotheses, each with its own unique predicted outcome, the experimenter is likely to have less of an allegiance to any one hypothesis. Therefore, results will be interpreted according to which hypothesis best accounts for the obtained findings.

Power and Generality of SI

Platt claims that SI is "*the* method of science and always has been" (p. 347). He chooses to give this method a special name because he is concerned that too many scientists have forgotten that this is "the" scientific method and they fail to teach their students this structured system of inference. The next section of this paper will attempt to point out difficulties inherent in the SI method that tend to complicate the process to the point that its usefulness is minimized for solving scientific problems. First, a critique of Platt's four steps of SI will be presented followed by an explication of other weaknesses in Platt's argument.

Critique of the Four Steps of SI

Step 1: Devising Alternative Hypotheses

The first step of SI is "devising alternative hypotheses." Although this may sound simple, there are several important problems. The most fundamental problem with this step is that it assumes that all problems addressed by the scientist are well defined; for example, there is a finite, specifiable set of possible alternative answers such that the correct answer is assured to be contained in this set. Platt fails to consider that hypotheses are proposed solutions to problems and therefore there is a prior step: the formulation of the scientific problem. Popper (1981) sees science as starting and ending with problems. He uses the following schematic:

$$P_1 \text{->} TS \text{->} EE \text{->} P_2$$

That is, an initial problem leads to a tentative solution, which leads to error eliminating tests, which lead to a reformulated problem. However, at times it may be the case that the scientist does not have sufficient background knowledge about a phenomenon or problem to formulate a problem in a way that it can be solved. One reason for this is that the background knowledge used to specify the problem may simply be wrong or incomplete in ways that fail to place the problem solution among the set of alternative hypotheses. Thus, all further inquiry is doomed to fail.

For example, Platt discusses this question from organic chemistry, "Do the bonds alternate in benzene or are they equivalent?" This problem is quite different from a currently unsolved problem in psychology, "What causes depression?" The first is a very specific, well-defined question with a finite set of alternative hypotheses (i.e., two), which renders the problem potentially amenable to SI. The latter question, on the other hand, is more vague and not well specified. Should guilt regarding one's sins be on the list of possible answers? If so, we then obtain a very large set of alternatives as various religions each define sin distinctly. Or, on the other hand, should deprivation be on the list because one has not sufficiently sampled sins? Should the quality of toilet training be on it? Residing near high-tension wires? Even if all possible candidates could be listed, the actual cause of depression might not be due to one factor but rather be found in the confluence or interaction of several factors and thus all permutations of all possible candidates would also need to be listed as alternatives. Popper saw this problem (Popper, 1981, p. 15) and argued that the refutation of competing hypotheses will never lead to a problem solution because "the number of *possibly* true theories remains infinite, at any time, and after any number of crucial tests." It is also likely that these more vague, open-ended questions may be quite common in young sciences (such as psychology) or at the frontiers of more established sciences and, thus, an enumeration of alternative hypotheses may not be possible in these cases (Hafner & Presswood, 1965).

Problem-solving strategies fall under two general types, algorithms and heuristics (Simon, 1979). An algorithm is a procedure that guarantees the solution of a problem, while heuristics are loose guidelines for problem solutions that may or may not lead to the solution of a problem. Platt seems to regard SI as an algorithm since he appears to assume that all problem statements can be defined in a way that makes it possible to enumerate a set of possibilities that contains the correct problem solution. Algorithms are rarely used in human problem solving and are highly inefficient because the closed set of alternative hypotheses may be very large (Simon, 1979). For example, the number of possible outcomes in chess is 10^{40} (Best, 1992). Therefore, it may be more accurate to conceptualize SI as a heuristic, for example, a strategy that may lead to the solution of a problem or may be used to solve only certain kinds of problems with particular characteristics (e.g., well-specified problems with only a few possible alternatives).

Platt might respond to the impossibility of enumerating all possible alternatives by arguing that it is necessary only to specify all "plausible" hypotheses. However, what will determine whether a hypothesis is plausible? It may be true that some portion of the possible alternatives are outlandish given current beliefs, but it is risky to rule these out antecedently because background presuppositions may be incorrect. An exhaustive list of plausible alternatives is best accomplished when scientists already have a good understanding of the phenomenon of interest. Again, this extensive background knowledge may not always be present at the frontiers of science or in young sciences such as psychology (Hafner & Presswood, 1965).

Moreover, antecedent theoretical preferences that vary across groups of scientists can also determine what is considered "plausible." For instance, when hypothesizing about the causes of depression, a cognitive psychologist may claim that particular kinds of distorted thought processes are the cause and he may spend a great deal of time and money investigating this hypothesis. However, a radical behaviorist may find this hypothesis implausible and not worth the intellectual effort and monetary resources. Thus, coming to an agreement on what is "plausible" may be difficult, as plausibility is based partly on background assumptions that vary across investigators.

Step 2: Devise Crucial Experiments

Platt states that the next step in SI consists of "Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses" (p. 347). The difficulty here is best captured by the Quine-Duhem problem (Gibson, 1982). Quine has argued that the scientist can never logically rule out a hypothesis even if experimental evidence is inconsistent with that hypothesis. Hypothesis testing utilizes a set of auxiliary hypotheses (Hempel, 1966). Therefore, an apparent falsification of a hypothesis could logically and plausibly be due to either the weakness of the central hypothesis or the weakness of any of the auxiliary hypotheses used to test the central hypothesis (e.g., poor measurement instruments, or a poor sample). Logically, the results of a crucial experiment cannot determine which of these ought to take the blame and be abandoned.

> What the scientist discovers in the case of an apparent falsification is an inconsistency; it is clear that some sort of revision of the scientist's belief set must occur, but it is much less clear which belief ought to be discarded. As Quine (1978) notes, [w]hen our system of beliefs supports our expectation of some event and that event does not occur, we have the problem of selecting certain of our interlocking beliefs for revision . . . [W]henever something expected fails to happen[,] we are called upon to go back and revise one or another of the beliefs that, taken together, had engendered the false expectation. (Gibson, 1982, p. 20)

The following example in Quine (1978, p. 103) illustrates his point:

> [C]onsider . . . the hypothesis that pure water under 760 millimeters boils at 100 degrees Celsius. Suppose that a quantity of what we take to be very pure water under a pressure of very nearly 760 is found to boil at 92 degrees Celsius. What conflict are not merely (a) the hypothesis and (b) the boiling at 92 degrees Celsius. Two further parties to the conflict immediately stand forth: the belief that (c) the water was pure and (d) the pressure was close enough to 760. Discarding the hypothesis is one way of maintaining consistency, but there are other ways. In the imagined case, of course, no scientist would reject the hypothesis, because it is such a highly accredited part of chemistry. Instead he would challenge (b), (c), or (d), each of which is open to question. In fact (b), (c), and (d) each rest in turn on some physics or chemistry, through the methods we use for determining what the temperature and pressure of a given liquid are, and when a liquid comes to a boil, and whether a liquid is pure water. So more than observation has entered into our acceptance of (b), (c), and (d). And even if we question no general beliefs underlying (b), (c), and (d), there could be some mistaken reading of dials and gauges. What has come to grief in our example is a whole family of beliefs.

While there may not be a clear choice of which belief to reject in the face of counter-evidence or an unexpected result (given that there is no set of rules that governs the process of rational belief revision), scientists assume, nevertheless, there are better or worse candidates for rejection and that certain pragmatic considerations can aid rational belief revision. Scientists generally will want to revise as few beliefs as possible, which will lead them to prefer those beliefs with the fewest implications and the least amount of support in their favor. For example, a law of physics that has survived attempted falsifications and possesses tremendous predictive and explanatory power should obviously not be high on anyone's list of candidates for rejection. Returning to Quine's example, in similar cases in the history of science, we can witness scientists exercising options similar to (b), (c), (d), and the assumption of having misread gauges and dials (see also Feyerabend, 1975). So the alternative hypothesis that conflicts with observation does not automatically face rejection, but as conflicting observations become more numerous across multiple experiments (not a single experiment as in SI), the alternative becomes an increasingly likely candidate for rejection. At what point rejection of the alternative should occur and how many anomalous observations can be ignored or explained by other means also, of course, requires a certain amount of judgment. However, Platt fails to see that replications (direct and indirect) may be necessary before an alternative is abandoned.

There is another reason why scientists should not discard a hypothesis even if a crucial experiment produces data inconsistent with it. It may be the case that examining possible problems with auxiliary hypotheses and making seemingly *ad hoc* adjustments to either the central hypothesis or auxiliary hypotheses may be a useful strategy for making scientific progress. Typically, making *ad hoc* adjustments is considered "unscientific" because it is simply an effort to save a hypothesis that is threatened by conflicting evidence (Hempel, 1966). However, Feyerabend (1988) argues that there are important historical examples in which making these *ad hoc* adjustments has been necessary in the progress of science. He cites as an example Newton's theory of colors. Newton argued that light consisted of rays that have a small lateral extension in space. Because the surfaces of mirrors are much rougher than the lateral extension of the rays, mirror images should not occur (if light consists of rays). However, mirror images do, obviously, occur. Newton responded to this contrary evidence by making the *ad hoc* hypothesis, "the reflection of a ray is effected, not by a single point of the reflecting body, but by some power of the body that is evenly diffused all over its surface." In this way, Newton corrected a discrepancy between theory and fact while retaining the usefulness of the theory (Feyerabend, p. 44, 1988).

Another difficulty with the second step is the notion "crucial" is not wholly an objective, logical matter but in part a subjective, social judgement. Not all scientists will agree that an experiment is actually a "crucial" experiment. For instance, a behavioral psychologist may wish to perform a crucial experiment in order to determine whether systematic desensitization or Freudian psychoanalysis is more effective for treating specific phobias. A study may then be set up in which a group of subjects are randomly assigned to sixteen sessions of either behavioral or psychodynamic therapy, with the dependent variable being how close subjects are willing to approach the feared object. Even if this study supported the greater efficacy of behavior therapy, a Freudian is not likely to see this single test as crucial. He is likely to argue several points, for example, more than sixteen sessions are needed for psychoanalytic treatment to be effective or that even if behavior therapy cured symptoms, there are still underlying disorders left untreated. Should the behavioral researcher go to the expense of providing full psychoanalytic treatment for these subjects (perhaps 500-600 sessions) and measuring other facets to detect possible symptom substitution? Therefore, even the issue of what exactly constitutes a "crucial" experiment is ambiguous.

Step 3: Conduct Crucial Experiments

Platt's third step is to "carry out the experiment as to get a clean result" (p. 347). Situations may arise in which the results of an experiment do not favor one hypothesis over the other. For instance, what if hypothesis 1 predicts a result of 10, while hypothesis 2 predicts a result of 14? Now, further assume that the obtained result is 12. What now can be said? Are both hypotheses ruled out? Are both retained? Are one or both of the hypotheses adjusted (e.g., modify hypothesis 2: "The result should be greater than 11")? Does the scientist retain the original hypotheses, but repeat the study after making some adjustment in the experimental procedure such as using different samples or measurement devices? Should the result be taken to indicate that there are additional alternative hypotheses that the scientist has failed to explicate?

Another difficulty with this step is that Platt assumes that research is conducted in a top-down fashion, meaning that we begin with hypotheses, test these hypotheses, see how the data turn out, and then, if necessary, adjust our hypotheses. What are ignored are situations where data are collected (possibly by accident) and then hypotheses are formed that simply summarize the data. For instance, serendipitous observations may lead to scientific discoveries without the prior formation of hypotheses. This occurred when the apparatus B. F. Skinner used to deliver food pellets to his rats malfunctioned (Skinner, 1956). He observed that the rate of responding of the rats increased dramatically even though the rats were no longer receiving food after pressing the lever (Skinner, 1956). This is what is now commonly known as an "extinction burst." In addition, at one point when Skinner was running low on rat food, he began reinforcing responses every minute instead of reinforcing every response (Skinner, 1956). As a result, Skinner noticed fairly stable rates of responding and was eventually able to describe the effects of different schedules of intermittent reinforcement (Skinner, 1956). In general, Skinner's version of the scientific method was to arrange controlled environments and then observe and record behavior in order to see if there were any regularities.

There is also the possibility that a hypothesis that is ruled out at one time may still be useful or even "correct." For example, the claim that the earth was in motion was a hotly debated issue in the early 1600's (Feyerabend, 1988). A "crucial experiment" was devised that involved dropping a stone from a tower to rule out one of these hypotheses (i.e., either the earth is moving or it is not). The outcome, of course, was that the stone landed at the base of the tower, which indicated to most that the earth was not moving (otherwise the stone should have fallen at some point other than directly below the point of release). However, the interpretation of the critical experiment ignored inertial velocity.

In fact, Galileo was able to use evidence from the stone demonstration as well as reason and persuasion to argue that the earth did, in fact, move (Feyerabend, 1988). To do this, he first had his reader imagine that they are on a moving ship. He then asked whether one would have to move their eyes in order to keep them focused on the ship's mast (or any other point on the ship). The answer is, of course, "no" because both the reader and the mast are moving along with the ship. The mast is moving, but the reader cannot detect this movement because the reader is also moving. Galileo argued that in the same way, as the earth moves, the reader, the stone, and the tower move along with it. Therefore, as the stone drops, the circular motion it should make (according to those not believing in the motion of the earth) will not be detected because the reader shares this motion with the stone, just as the ship's mast does not appear to move when both the reader and it are standing on the moving ship.

Step 4: Recycle

According to Platt, the final step of SI involves "recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain" (p. 347). There are several aspects of this step that are unclear. First, it is not clear what exactly "subhypotheses" or "sequential hypotheses" are. A definition of these terms would be helpful. Second, it is not clear how one is to construct these subhypotheses or sequential hypotheses. Are these actually supposed to be deduced in some logical manner, and, if so, what is the nature of this sequence? Are these new hypotheses to new problems? Finally, it is unclear why one would have to do this if step 1 was followed. If an exhaustive list of alternative hypotheses was already made, new ones, in principle, should not arise.

Problematic Historiography

The accuracy of the historical analyses in SI is also problematic. Platt at least indirectly claims that SI is *the* method of science and always has been (p. 347). Platt also makes another historical claim when he asserts that what distinguishes the successful from the less successful sciences is that the successful sciences have in the past used SI while the unsuccessful sciences have not. We note first that Platt does not provide sufficient evidence from the historical record to support either of these claims. That is, he provides no general analysis to show that the successful sciences have always used SI or even that they have more consistently used SI than the unsuccessful sciences. We also note that Platt asserts something quite different than other historians of science such as Thomas Kuhn (1970) about what a historical study of science reveals about what is the proper scientific method. In this section, we will provide a more detailed analysis of some of the historical case studies Platt cites in SI. In addition, historical examples of scientific methods other than SI that lead to significant scientific achievements also will be discussed.

Did the Studies Cited by Platt Actually Utilize SI?

Platt discusses several examples of studies from the fields of molecular biology and high-energy physics that he claims used SI methodology. His discussions of these examples are quite cursory. Platt sometimes asserts that a particular scientist (for example, Roentgen or Pasteur) used the SI method, but does not provide sufficient supportive details demonstrating how these individuals actually utilized each step in the scientific discovery. At least one or two detailed accounts of the explicit use of SI would provide more persuasive evidence for the claim that these scientists actually used SI.

Furthermore, it is not clear whether the examples Platt cites are actually examples of SI. For instance, Platt claims Lederberg (1959) used language to describe his research that is consistent with SI (p. 348). The scientific problem of Lederberg's paper is to outline an elective theory of antibody formation. This elective theory is an alternative to the instructive theory, so if this paper were consistent with SI language, it seems that the paper would describe crucial experiments in order to rule out one of these alternatives.

In outlining this elective theory, Lederberg states nine propositions of antibody formation. These propositions are not alternative hypotheses. Rather they are a series of premises on which the truth of this elective theory depends. Platt indicates that Lederberg is claiming that the list of propositions is "subject to denial" (p. 348). A more careful reading of Lederberg shows that this quote is taken out of context because: 1) the quote only applies to four of the propositions, not all nine, and 2) the quote actually reads " . . . however, they [the first four propositions] might well be subject to denial or modification *without impairing the validity of the elective approach*" (emphasis added; p. 1649). Furthermore, "the last four propositions are stated to account for the general features of antibody formation in cellular terms *and may be equally applicable to instructive and elective theories*" (italics added; p. 1649). Therefore, only one of these propositions applies *exclusively* to this elective theory and no crucial experiments are proposed regarding this proposition. Therefore, the propositions "subject to denial" are not critical for supporting or refuting the elective theory in favor of the alternative instructive theory, which would presumably be the purpose of utilizing SI.

In a section on molecular biology, Platt refers to a study by Meselson and Stahl (1958), citing it as an example of a crucial experiment designed to determine if the strands in DNA separate or remain in tact when a cell divides. However, no set of alternative hypotheses was enumerated in this paper. Certain passages hint at a set of alternatives, but these are not specified. For instance, "Hypotheses for the mechanism of DNA replication differ in predictions they make concerning the distribution among progeny molecules of atoms derived from parental molecules" (p. 671). However, these predictions are not specified. Also, the statement, "This suggested to Watson and Crick a definite and structurally plausible hypothesis for the duplication of the DNA molecule. According to this idea, the two chains separate, exposing the hydrogen-bonding sites of the bases" (p. 677), specifies one hypothesis, but not an exhaustive set of alternatives. Therefore, this study was designed to simply test the validity of this one hypothesis. However, had the study refuted the hypothesis, it would not have supported some specified alternative.

Also in the molecular biology section, Platt discusses a study by Benzer (1959) concerning the structure of the "genetic map." Again, no set of alternative hypotheses was delineated. In fact the author states, "Our objective is to examine the topology of the structure in phage T4 that controls its ability to multiply in K, and specifically to make a rigorous test of the notion that the structure is linear" (p. 1610). Thus, the experiment was apparently designed to test the validity of one hypothesis (i.e., the structure is linear), not to rule out one hypothesis in favor of another alternative hypothesis.

In addition, Platt's account of this study is incomplete. Platt states, "Seymour Benzer showed that his hundreds of fine microgenetic experiments on bacteria would fit only the mathematical matrix for the one-dimensional case" (p. 348). A careful reading of the original article, however, shows that Benzer stated that "a simple linear model suffices to account for the data" (p. 1619). In addition, he also stated that an alternative account, that the structure is two-dimensional (branched), cannot be ruled out. For example, he stated, "It is in the nature of the present analysis that the existence of complex situations cannot be disproved" (p. 1619), and "The possibilities of branches within the structure is not necessarily excluded" (p. 1619). Therefore, although Benzer appropriately favored the more parsimonious explanation, the experiment conducted was not an example of SI because a set of alternatives was not specified nor was it a crucial experiment that functioned to rule out other alternatives. Moreover, even if we grant this was a crucial experiment, it did not produce "clean" results. No suggestions for recycling (step 4) were provided, except for perhaps that further studies should attempt to exclude complex structures. Overall, we claim, contra Platt, that there is no good evidence that any of the SI steps were followed.

In his discussion of the use of SI in high-energy physics, Platt cites a study conducted by Garwin, Lederman, and Weinrich (1957). The purpose of their study was to determine if "fundamental particles conserve mirror-symmetry or 'parity' in certain reactions, or do they not?" (Platt, p. 349). It does appear from Garwin et al., (1957) that their study was a crucial experiment, attempting to answer this question and that it did rule out parity conservation. The problem is that the alternative hypothesis is simply "or is not the case." Therefore, it may be that one hypothesis was ruled out, but in favor of what specific alternative? Alternative hypotheses (step 1) were never enumerated.

Platt also praises Roentgen's work on the description of the properties of X-rays (Watson, 1945) as an exemplary of inductive inference. First, the actual discovery of X-rays, which seems at least as significant and interesting as describing their properties, was not achieved through the use of SI. Second, Platt states Roentgen's work is an example of *inductive* inference, not SI. Thus, it is unclear why Platt attempts to use this historical case to support the thesis that SI was used in this successful scientific episode. In fact, Roentgen's paper does not specify alternative hypotheses that were tested through crucial experiments. Some of the experiments described had an "exploratory—what would happen if?" attitude—described above as a method Skinner preferred—more than an exemplification of SI. For instance, Roentgen attempted to see if X-rays could penetrate various substances such as aluminum, water, paper, and wood. He found that X-rays penetrated all of these substances, but that the degree of transparency depended on density of the material. He also attempted to refract X-rays through a prism and deflect them with a magnet. There were neither alternative hypotheses devised nor crucial experiments conducted. Roentgen simply tinkered with these newly discovered X-rays to see what they could and could not do and from this abstracted what he saw as an admittedly nonexhaustive list of the properties of X-rays.

Feyerabend (1975) has claimed that scientists write revisionary accounts that provide cleaner, more logical versions of what actually occurred and that conform to what is commonly taken to be good scientific practice. Hafner and Presswood (1965) describe just such a case in physics where " . . . the principle of SI failed." (p. 509). They describe the history of the "universal Fermi interaction," which accounts for beta decay (the release of electrons or positrons and an antineutrino or neutrino by the nucleus of atoms in radiating matter such as uranium). They present an idealized form of this line of research near the end of their paper (p. 509), which could be interpreted as an instance of SI. However, the majority of the paper is devoted to demonstrating the actual series of events that led to "V-A theory" of the universal Fermi interaction, a series of events that is much different from the actual steps of SI. For instance, they describe a process characterized by several events not consistent with SI such as: 1) the conduction of experiments in the absence of hypotheses to be tested; 2) theory formation *following* experimental findings; 3) results from crucial experiments being wrong and misleading; and 4) the results of experiments being ignored because they did not fit a theoretical framework. They conclude that, " . . . SI is an idealized scheme to which scientific developments seldom conform" (p. 503).

Other Historical Cases—Did Galileo, Darwin, Lyell, or Newton Use SI?

Galileo. Galileo made numerous important contributions to the study of motion and astronomy and is considered one of the founders of science. His contributions included descriptions of uniformly accelerate straight motion (i.e., the times-square law of distance in fall), projectile motions, and the pendulum law (Drake, 1990). He also described the moons of Jupiter, the surface of the moon, the phases of Venus, and sunspots (Ravetz, 1990).

The scientific method used by Galileo was clearly not SI. His form of the scientific method will first be described and then illustrated with specific historical examples. Galileo believed the purpose of science was to establish general truths about motion (Gower, 1997). Therefore, he was committed to the view that the scientific method should involve the use of deductive reasoning, because he argued that inductive reasoning can never lead to universal claim. Galileo thought that the scientist can never observe every possible instance of a particular phenomenon and, thus, induction would always be inconclusive (i.e., the "white swan" example, Gower, 1997).

Galileo's form of deductive reasoning was strongly influenced by that of Euclid and Archimedes (Drake, 1990). This method began with intuitively appealing and generally uncontroversial premises, what are sometimes referred to as Archimedian points (Gower, 1997; Snow, 1975). These axioms then form the foundation upon which theorems about the world can be logically derived (Drake, 1990; Snow, 1975). Galileo made use of experiments as a means for demonstrating that these premises, or axioms, characterize the phenomenon of interest.

An example of this process is Galileo's formation of the times-squared rule of distance in free fall. First, Galileo developed an axiom describing naturally accelerated motion. The resulting axiom defined uniformly accelerated motion as motion in which the velocity of a free falling object is proportional to the time that it has been falling (Gower, 1997). From these axioms, Galileo deduced consequences about relations between time, velocity, and distance. Galileo then performed experiments in which he measured the distance a ball rolled down an inclined plane in equal intervals of time in order to demonstrate these proposed relations (Gower, 1997). What resulted was the times-squared rule, which could also be mathematically derived as a theorem from the definition of naturally accelerated motion described above (Gower, 1997).

Galileo also employed the use of what he called "thought experiments." He argued that thought experiments should be an essential component of the scientific method, because he argued that the results of a single experiment depended upon many accidental properties of particular bodies and events and as such could produce unreliable results (Drake, 1990; Gower, 1997). Thought experiments, on the other hand, did not contain any of the flaws of real experiments and in addition were accessible to all, thus having the persuasive force of common observation (Gower, 1997). Therefore, thought experiments were used as a means for demonstrating that free fall is a phenomenon of uniformly accelerated motion. His clever use of thought experiments lead readers to the conclusion that his definition of naturally accelerated motion was the only reasonable definition. For example, Galileo's definition of naturally accelerated motion states that the velocity of a free falling object is proportional to the time it has been falling. However, an Aristotelian account states that the velocity of a free falling object is proportional to its weight (Drake, 1990). The following thought experiment shows that Aristotle's account contains an inherent contradiction, and therefore, cannot be a true account of nature (Drake, 1990).

Galileo asked his reader to suppose that he had two stones, one being twice the weight of the other. According to Aristotle, the heavier one will fall twice as fast as the lighter one. Next the reader is asked to assign the stones speeds of 2 and 4. Now, say we tie the two stones together and drop them. The Aristotelian view predicts that the lighter stone will slow the descent of the heavier stone. Now we glue the two stones together. We have created a stone that is of greater weight than required for a speed of 4, but should fall at a speed of less than 4 according to the Aristotelian view. Therefore, a contradiction is present because the conjoined stones cannot both fall faster and slower than 4 units. Since Galileo's account contains no such contradiction, Galileo argues that it must be the correct account.

It should be noted that the above example resembles SI. Furthermore, Galileo's "Dialogue on the Two Chief World Systems" (Galileo, 1632/1962) does admittedly pit alternative views against each other (i.e., Ptolemaic and Copernican models of the solar system). However, these examples do not seem to represent SI as outlined by Platt. First, although he does mention one alternative hypothesis (i.e., the Ptolemaic model), there is no formalized list of all possible alternative hypotheses (step 1). Second, there does not appear to be a process of conducting crucial experiments in order to "clip branches off the inductive tree" and come to the "final answer," primarily because there are only two alternatives. Finally, there seems to be no process of recycling procedures or making subhypotheses (step 4).

To summarize then, Galileo made use of axiomatic deductive reasoning as his preferred method of science. This method starts with premises or axioms that are considered to be generally accepted and uncontroversial. Then, these axioms could be used to deduce, through the use of mathematics or reason, further knowledge (such as relations between time, velocity, and distance of free falling objects). Experimentation, both in the laboratory and in thought, was used to demonstrate knowledge deduced from axioms. Mathematical generalizations, such as the times-square law, may then result (Snow, 1975). It is this unique blend of experimentation and mathematics that is considered to be Galileo's primary contribution to the development of the scientific method (Gower, 1997).

Darwin. The work of Charles Darwin represents another example of a scientist making significant contributions without the use of SI. Darwin's method seems to have initially involved the formation of what Ghiselin (1969) calls "hypothetico-deductive systems." This involves reasoning from a series of general postulates to a series of specific theorems that can be used to explain a phenomenon and predict various outcomes that can be tested experientially (Johnston & Pennypacker, 1993). These hypothetico-deductive systems were generally based on a great deal of observation and reasoning, as was the case with his theory of natural selection that was developed following his famous 5-year journey on the *Beagle* (Ghiselin, 1969; Himmelfarb, 1967). Darwin admits the importance of reasoning when he states that, "... the *Origin of Species* is one long argument from the beginning to the end ... " and that, "no one could have written it without having some power of reasoning" (p. 82). In addition, his theory concerning the formation of coral reefs seems to be almost solely the product of reasoning, given that he formed this theory prior to ever observing a coral reef (Darwin, 1888, p. 58). This theory was, instead, derived from discoveries made while on the *Beagle* that masses of land can rise and subside (Ghiselin, 1969).

Once these hypothetico-deductive systems were developed, Darwin sought to make specific, verifiable predictions that logically followed from the series of premises present in the system (Ghiselin, 1969). If the prediction matched observation, the system was supported, if not, the system had to be altered. Darwin, however, was not interested in simply collecting a series of observations that were consistent with his system, but was also interested in observations that refuted his system (Darwin, 1888, pp. 71 & 83). Therefore, Darwin's method seems to be a mix of detailed observation, powerful reasoning, theory construction, deducing predictions from the theory, testing these predictions, and altering the theoretical system if necessary.

Although Darwin's emphasis on making predictions and attempting to refute them through observation may resemble SI, it differs in important ways. For instance, there appears to be no process of forming lists of alternative hypotheses. Also, no crucial experiments are set up and conducted. In the areas of geology and natural selection, experiments cannot even be done because one cannot manipulate the phenomenon of interest, making these areas unamenable to the SI method. Darwin's method seems to be more of a process of forming theories, making predictions, observing whether the prediction is accurate based on observations, and then altering theories if the observations do not match the predictions. There does not seem to be a process of pitting hypotheses against each other, throwing the refuted ones out, and then arriving at the "best" one.

Lyell. Charles Lyell provides a somewhat more obscure, but equally significant example of the formation of a scientific method for the science of geology that is also inconsistent with SI. Lyell rejected hypothesizing primarily because one could not control or manipulate the phenomenon to be studied in geological sciences (Laudan, 1990). Therefore, SI could not be conducted because crucial experiments could not be set up. Instead, Lyell proposed a method of true causes in which causes of phenomenon are postulated. These causes must satisfy two conditions according to Lyell: 1) these must be known to exist and 2) these must be known to be adequate to produce the effects. In geology, Lyell argued that one could satisfy these conditions by observing postulated causes and observing these causes producing the postulated effect (Laudan, 1990). Although SI does not apply to nonexperimental sciences, Lyell represents an example of how important scientific achievements can be made without the use of SI.

Newton. Interestingly, Platt admits that Newton's extensive scientific contributions were made without the use of SI, but downplays these as "rare and individual achievements that stand outside any rule or method" (p. 351). However, it seems inappropriate to label Newton's achievements (or any of the other scientists mentioned) as simply a "rare and individual achievement" that stands outside the use of a particular method.

Thus, we have illustrated how at least four other significant scientific contributions all were made without reliance on SI. This further calls into question of whether SI is the method of the successful sciences.

Is There One Scientific Method?

It could, in fact, be argued that the unique achievements made by the individuals mentioned above, using a wide range of methods, may actually support the notion that adherence to any single rule or method is unnecessary, perhaps even harmful. Some authors, such as the philosopher of science Paul Feyerabend (1988), would argue that the idea that scientific progress has been made by adherence to a specific method is historically inaccurate:

> The idea of a method that contains firm, unchanging, and absolutely binding principles for conducting the business of science meets considerable difficulty when confronted with the results of historical research. We find, then, that there is not a single rule, however plausible, and however firmly grounded in

> epistemology, that is not violated at some time or other. It becomes evident that such violations are not accidental events, they are not results of insufficient knowledge or of inattention which might have been avoided. On the contrary, we see that they are necessary for progress. (p. 14)

Furthermore, failure to strictly adhere to a specific method may be how scientific progress is actually made. Feyerabend (1988) claims that many scientific achievements such as the Copernican revolution, atomism, and wave theory of light, "occurred only because some thinkers either *decided* not to be bound by certain 'obvious' methodological rules, or because they *unwittingly broke* them" (p. 14). SI would certainly be an example of the "obvious methodical rules" or "binding principles" Feyerabend is referring to. Therefore, to claim that SI is *the* method of scientific inquiry may not only be historically inaccurate, but promoting a form of science that could stifle creativity and progress.

In addition, some cautionary guides can be drawn from our analysis of SI. First, generalizations from one subject domain regarding what is a good method to another should be made cautiously. None of Platt's examples were from the social sciences. T. Nickles (personal communication, March 21, 1998) has argued that any method of discovery makes assumptions about both the cognitive/perceptual abilities of the scientist as well as the nature of the phenomenon being investigated. Like a lever, the method must appropriately connect at both ends. As phenomena being investigated vary, the lever may or may not connect at this end. Second, although falsification has considerable advantages, particularly, in that it can offset confirmation biases, it is not a mechanical method or a heuristic that will necessarily lead to the solution of scientific problems. The Quine-Duhem problem illustrates the logical inconclusiveness of falsificationism and how a scientist's judgement comes into play. Gower (1997, p. 241) states, "But we do not so readily acknowledge the wide scope of these [experimental] skills. Imagination and ingenuity are clearly important; so are interpretative and manipulative capacities." Third, problem formulation and specification (particularly with the ontic assumptions and putative background knowledge that will necessarily be involved in any problem statement) is a critical step in science. Unfortunately, there is no mechanical method that assures the problem is stated in a way that it can be solved. Fourth, what counts as a problematic *ad hoc* adjustment in the face of an apparent falsification is actually a complex matter. Philosopher of science Imre Lakatos (1976) has suggested that the judgement of whether *ad hoc* adjustments or efforts to save a hypothesis are problematic cannot be made in a single experiment, but rather the unit needs to be a research program. If the research program fails to attribute the anomaly to a falsification of the major hypothesis under test and instead places the blame for the prediction failure to an auxiliary hypothesis, and this modified auxiliary hypothesis results in future new predictions and corroborations, then Lakatos argues that this is an acceptable scientific move. Fifth, the notion of a "critical experiment" is again not simply a mechanical, logical designation. It involves a sociological, rhetorical component in that groups of people with differing commitments need to be persuaded. Any scientist attempting to conduct critical experiments needs to be aware of the sometimes daunting persuasive task before him or her (Gross, 1996).

Summary and Conclusions

Platt's SI has had a large influence on the social sciences as well as science in general. The purpose of this paper was to argue that SI may not be as simple as Platt would lead us to believe. Each step is fraught with problems that make SI difficult to actually implement. Furthermore, it is unclear whether the historical cases Platt cites as good examples of SI actually used SI. In addition, other historical examples were presented to show that scientific progress does not require the use of SI. Overall, there does not appear to be any evidence to suggest that SI accounts for what progress has occurred in science, what is currently being done in science, or is a method that will inevitably lead to scientific progress in the future. These problems may explain why despite SI being frequently touted as a regulative scientific method, it is infrequently used.

References

Benzer, S. (1959). On the topology of the genetic fine structure. *Proceedings of the National Academy of Sciences, USA, 45,* 1607-1620.

Best, J. B. (1992). *Cognitive psychology* (3rd ed.). St. Paul, MN: West.

Bordens, K. S., & Abbott, B. B. (1991). *Research design and methods: A process approach.* Mountain View, CA: Mayfield Publishing Company.

Borkovec, T. D. (1997). On the need for a basic science approach to psychotherapy research. *Psychological Science, 8,* 145-147.

Chamberlin, T. C. (1965). The method of multiple working hypotheses. *Science, 148*; 754-759.

Darwin, F. (1888). *The life and letters of Charles Darwin, including an autobiographical chapter* (Vols. 1-2). New York: D. Appleton and Company.

Dixon, J. A. (1998). Developmental ordering, scale types, and strong inference. *Developmental Psychology, 34,* 131-145.

Drake, S. (1990). *Galileo: Pioneer scientist.* Toronto: University of Toronto Press.

Feyerabend, P. (1975). *Against method: Outline of an anarchistic theory of knowledge.* London: Humanities Press.

Feyerabend, P. (1988). *Against method* (2nd ed.). London: Verso.

Folger, R. (1989). Significance tests and duplicity of binary decisions. *Psychological Bulletin, 106,* 155-160.

Galileo, G. (1962). *Dialogue concerning the two chief world systems—Ptolemaic & Copernican.* (S. Drake, Trans.). Berkeley: University of California Press. (Original work published 1632)

Garwin, R. L., Lederman, L. M., & Weinrich, M. (1957). Observations of the failure of conservation of parity and charge conjugation in meson decays: The magnetic moment of the free muon. *Physical Review, 105,* 1415-1417.

Ghiselin, M. T. (1969). *The triumph of the Darwinian method.* Berkeley, CA: University of California Press.

Ghosh, B. N. (1984). *Scientific method and social research.* New Delhi: Sterling Publishers.

Gibson, R. F. (1982). *The philosophy of W. V. Quine: An expository essay.* Tampa, FL: University Presses of Florida.

Gower, B. (1997). *Scientific method: A historical and philosophical introduction.* London: Routledge.

Gross, A. G. (1996). *The rhetoric of science.* Cambridge: Harvard University Press.

Hafner, E. M., & Presswood, S. (1965). Strong inference and weak interactions. *Science, 149,* 503-510.

Hempel, C. G. (1966). *Philosophy of natural science.* New York: Prentice-Hall.

Hertzog, C. (1994). Repeated measures analysis in developmental research: What our ANOVA text didn't tell us. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. 187-222). Hillsdale, NJ: Lawrence Erlbaum Associates.

Himmelfarb, G. (1967). *Darwin and the Darwinian revolution.* Gloucester, MA: Peter Smith.

Johnston, J. M., & Pennypacker, H. S. (1993). *Strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kendall, P. C., & Hammen, C. (1995). *Abnormal psychology* (2nd ed.). Boston: Houghton Mifflin Company.

Kuhn, T. (1970). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lakatos, I. (1976). History of science and its rational reconstruction. In C. Howson (Ed.), *Method and appraisal in the physical sciences* (pp. 1-39). Cambridge: Cambridge University Press.

Laudan, R. (1990). The history of geology, 1780-1840. In R. C. Olby, G. N. Cantor, J. R. R. Christie, & M. J. S. Hodge, (Eds.), *Companion to the history of modern science* (pp. 314-325). London: Routledge.

Lederberg, J. (1959). Genes and antibodies. *Science, 129,* 1649-1653.

Mackenzie, B. (1984). Explaining race differences in IQ: The logic, the methodology, and the evidence. *American Psychologist, 39,* 1214-1229.

Mahoney, M. J. (1980). *Abnormal psychology: Perspectives on human variance.* New York: Harper & Row.

Massaro, D. W., & Cowan, N. (1993). Information processing models: Microscopes of the mind. *Annual Review of Psychology, 44,* 383-425.

Meselson, M., & Stahl, F. W. (1958). The replication of DNA in Escherichia coli. *Proceedings of the National Academy of Sciences, USA, 44,* 671-682.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice-Hall.

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. (1981). Conjectural knowledge: My solution to the problem of induction. In K. R. Popper (Ed.), *Objective Knowledge* (pp. 1-31). London: Oxford.

Quine, W. V., & Ullian, J. S. (1978). *The web of belief* (2nd ed.). New York: Random House.

Rathus, S. A., & Nevid, J. S. (1991). *Abnormal psychology*. Englewood Cliffs, NJ: Prentice Hall.

Ravetz, J. R. (1990). The Copernican revolution. In R. C. Olby, G. N. Cantor, J. R. R. Christie, & M. J. S. Hodge, (Eds.), *Companion to the history of modern science* (pp. 201-216). London: Routledge.

Roberts, K. H., & Glick, W. (1981). The job characteristics approach to task design: A critical review. *Journal of Applied Psychology, 66,* 193-217.

Simon, H. A. (1979). Information processing models of cognition. *Annual Review of Psychology, 30,* 363-396.

Skinner, B. F. (1956). A case history in scientific method. *American Psychologist, 11,* 221-233.

Smith, B., & Sechrest, L. (1991). Treatment of aptitude X treatment interactions. *Journal of Consulting and Clinical Psychology, 59,* 233-244.

Snow, A. J. (1975). *Matter and gravity in Newton's physical philosophy*. London: Oxford.

Watson, E. C. (1945). The discovery of X-rays. *American Journal of Physics, 13,* 281-291.

Zimmerman, J., & Zimmerman, E. H. (1976). Towards humanizing education: Classroom behavior management revisited. *The Psychological Record, 26,* 387-397.