Title

Are Scientific Theories that Predict Dat a Mor e Believable than Theories that Retrospectively Explain Data ? A Psychological Investigation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 16(0)

Author
Chinn, Clark A.

Publication Date
1994

Peer reviewed

Are Scientific Theories that Predict Data More Believable than Theories that Retrospectively Explain Data? A Psychological Investigation

Clark A. Chinn
Center for the Study of Reading
University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, IL 61820
(217) 333-2552
chinn@vmd.cso.uiuc.edu

Abstract

Philosophers have disagreed about whether theories that make successful predictions are more believable than theories that merely explain data that have already been discovered. Predictivists believe that theories that make successful predictions have an edge over theories that offer only retrospective explanations of the same data. Nonpredictivists maintain that whether a theory predicts data or explains data retrospectively is irrelevant to the believability of the theory. The purpose of this paper is to report on three psychological experiments designed to determine whether undergraduates behave as predictivists or nonpredictivists when they evaluate theories. Results indicate that subjects behaved as nonpredictivists when one theory predicted a body of data and a second theory was devised later to explain the same data retrospectively. However, subjects behaved as predictivists in the situation in which a theory retreated in the face of anomalous data by adding an auxiliary hypothesis; for instance, theories that predicted data by adding the necessary auxiliary hypotheses before the data came in were more believable than theories that added the auxiliary hypothesis in reaction to the data. These results suggest that cognitive models of theory choice that assume that people are nonpredictivists may require modification.

An unresolved issue in the philosophy of science is whether theories that make successful predictions are, other things being equal, more believable than theories that merely explain data that are already known. According to the *predictivist* position (Lakatos, 1970; Maher, 1990, 1992, 1993; Popper, 1959), a theory that makes successful empirical predictions about new data is more credible than a theory that explains the same data after the fact. According to the opposite position (Brinberg, Lynch, & Sawyer, 1992; Horwich, 1982; Howson, 1991; Howson & Franklin, 1991; Keynes, 1921; Thagard, 1992), which I will call the *nonpredictivist* position, a theory that predicts new data is no more believable than a theory that retrospectively explains the same data. To a predictivist, the timing of hypotheses matters; hypotheses that predate (and predict) data are superior to hypotheses that merely postdate data. To a nonpredictivist, the timing of a theory is irrelevant; a theory's credibility is determined solely by how well the theory explains a body of data, regardless of whether or not the theory predicted some subset of the data.

The difference between predictivists and nonpredictivists is clearly seen in their interpretation of why scientists became convinced that Mendeleev's periodic table was correct. According to the predictivist thesis, Mendeleev's periodic table gained credibility because it successfully predicted the existence and properties of three as-yet-undiscovered elements (Maher, 1993). According to the nonpredictivist thesis, the periodic table was believable just because it provided a good account of the chemical data, and it would have been just as believable if it had been devised after these three elements had already been discovered (Howson & Franklin, 1991).

Many models of theory choice in cognitive science are based on a nonpredictivist position. For instance, Thagard (1992) has developed a connectionist model of theory choice called ECHO that chooses between rival theories on the basis of which theory explains a body of data most coherently, irrespective of whether the hypotheses of the rival theories predated or postdated the data. Although originally developed as a model of scientific revolutions in the history of science, ECHO has also been applied as a model of theory choice in nonscientists (e.g., Schank & Ranney, 1992). Psychologists who have discussed factors that influence theory choice seldom take the timing of hypotheses into account (e.g., Chinn & Brewer, 1993).

The purpose of this paper is to examine the issue of whether people (nonscientists) are predictivists or nonpredictivists when they evaluate theories. If people operate as predictivists, models of human theory choice should take the timing of theoretical hypotheses into account. If people operate as nonpredictivists, then models of human theory choice need not incorporate information about the timing of hypotheses.

Pilot Study

Method

To obtain a baseline measure of whether people tend to be predictivists or nonpredictivists, 30 undergraduates read a text of about 330 words describing two fictional theories that were called N. F. Theory and Brans-Dicke Theory. According to the text, N. F. Theory was developed in the 1930s to account for certain reactions involving radioactive decay. The theory was said to have predicted the emission of a massless, electrically neutral particle called the neutrino, which would be extremely difficult to detect. The text stated that other scientists scoffed at N. F. Theory because of its prediction of an undetectable particle. But in the 1950s, newly developed equipment enabled scientists to detect the neutrino, thus providing striking vindication for N. F. Theory. The text went on to state that in the 1980s a rival theory, Brans-Dicke theory, was developed by two scientists who examined all of the radioactivity phenomena explained by N. F. Theory and then developed a new theory that can explain all of these same phenomena, including the emission of neutrinos. According to the text, both N. F. Theory and Brans-Dicke Theory were consistent with all known data; the theories do make some different predictions, but these predictions can be tested only with the help of particle accelerators larger than we have today.

Half of the subjects read an additional paragraph of 75 words that made the prediction issue explicit. The paragraph explained that some scientists believe that N. F. Theory should get credit for making a bold prediction, whereas other scientists believe that it doesn't matter which theory made the initial prediction.

After reading the text, subjects rated the extent to which they believed each theory was correct on a 0 to 10 scale, and they wrote explanations for their ratings.

Hypotheses

If predictivism is correct, subjects should rate N. F. Theory higher because of its bold prediction. If nonpredictivism is correct, subjects should give N. F. Theory and Brans-Dicke Theory equal ratings.

Results

The results of this pilot study gave clear support to the nonpredictivist position. Almost all subjects (28 of 30) gave N. F. Theory and Brans-Dicke Theory equal ratings. It did not matter whether subjects read the paragraph that explicitly raised the issue of predictiveness; 14 of 15 subjects in each condition gave the two theories equal ratings.

For the configuration of theories and evidence used in this study, the predictivist hypothesis was contradicted.

Experiment 1

In the pilot study just described, the nonpredictivist thesis was supported. However, the situation used in the pilot study was atypical of actual science in one crucial respect. Subjects read about static theories, theories that were fully developed from the outset and were never changed thereafter. Actual science is dynamic, with theories undergoing constant revision to accommodate new data. Perhaps predictivism prevails under more dynamic circumstances.

As an illustration, consider this situation. Theory Y explains a body of data D, but new data D1 are discovered that contradict Theory Y as it currently stands. To explain D1, Theory Y must introduce an auxiliary hypothesis Y1. {Y + Y1} thus explains D1 retrospectively. Now, suppose that Theory Y had already included the auxiliary hypothesis Y1 prior to the discovery of D1. In this case, {Y + Y1} would have predicted D1.

A predictivist would argue that {Y + Y1} is more believable if Y1 is added prior to the discovery of D1 than if Y1 is added after the discovery of D1. However, because nonpredictivists deny that the timing of hypotheses influences the believability of those hypotheses, nonpredictivists should maintain that {Y + Y1} explains {D + D1} just as well when Y1 was added after the discovery of D1 as when Y1 was added prior to the discovery of D1. Experiment 1 was designed to test these two positions.

Method

Subjects. Seventy-six undergraduates enrolled in introductory educational psychology courses participated in the experiment.

Materials. Subjects read passages of 475 to 525 words. The passages began by describing two theories designed to explain why cool weather often causes plants to stop growing even after the weather warms up again. The *rubisco* theory attributed the halted growth to deactivation of the enzyme rubisco activase; the *carboxydismutase* theory attributed the suspended growth to deactivation of a different enzyme, carboxydismutase.

All subjects then read data in which an independent team of scientists found depressed levels of rubisco activase in chilled tomatoes but were unable to determine the levels of carboxydismutase; these results were consistent with the rubisco theory but potentially inconsistent with the carboxydismutase theory.

Next the subjects read about how proponents of each theory responded to this piece of data. The rubisco theory had predicted depressed levels of rubisco activase and therefore counted the data as a successful prediction, but the carboxydismutase theory required an auxiliary hypothesis to explain why levels of rubisco activase were depressed.

Finally, subjects rated their belief in each theory on a 0 to 10 scale, and they explained their ratings.

Table 1. Experimental Conditions in Experiment 1.

	Predictive	Reactive
Exception-barring auxiliary hypothesis	Rubisco theory predicts tomato data. Carboxydismutase theory says in advance that tomatoes are exceptions.	Rubisco theory predicts tomato data. Carboxydismutase theory says post hoc that tomatoes are exceptions.
Revised-pathway auxiliary hypothesis	Rubisco theory predicts tomato data. Carboxydismutase theory says in advance that deactivating carboxydismutase has the side effect of deactivating rubisco activase.	Rubisco theory predicts tomato data. Carboxydismutase theory says post hoc that deactivating carboxydismutase has the side effect of deactivating rubisco activase.

Design. Two factors were crossed in a 2 x 2 design. One factor was the type of auxiliary hypothesis used by the carboxydismutase theory. The *exception-barring auxiliary hypothesis* asserted that tomatoes were an exception, that the carboxydismutase theory did not apply to tomatoes. Exception-barring hypotheses are an extreme type of ad hoc response to anomalous data (Lakatos, 1976). The *revised-pathway auxiliary hypothesis* explained away the depressed level of rubisco activase by altering the biochemical pathway assumed by the carboxydismutase theory. Carboxydismutase theorists said that the deactivation of carboxydismutase produced the side-effect of deactivating rubisco activase, but that the deactivation of rubisco activase did not play a causal role in stopping plant growth.

The second factor was whether the auxiliary hypothesis in the carboxydismutase theory was *predictive* or *reactive*. As an example, consider the exception-barring auxiliary hypothesis. In the predictive condition, carboxydismutase theorists asserted from the very start, prior to the experiment with chilled tomatoes, that tomatoes were an exception to the theory. In the reactive condition, carboxydismutase theorists asserted that tomatoes were an exception to the theory only after seeing the data from the chilled tomatoes; here the exception-barring appears to be an ad hoc reaction to the data. The revised-pathway auxiliary hypothesis was made either predictive or reactive in a similar way.

Table 1 presents a synopsis of the four conditions in the 2x2 design.

Hypotheses

Notice that for each type of auxiliary hypothesis, the final configuration of data and theory is the same in the predictive and reactive conditions. In each case, the carboxydismutase theory plus one auxiliary hypothesis is used to explain the same piece of data. The only difference is the timing of the introduction of the auxiliary hypothesis in the carboxydismutase theory. Thus, the nonpredictivist should predict that the predictive and reactive versions of the theory are equally believable. The predictivist would expect the reactive auxiliary hypothesis to weaken belief in the theory, compared with the predictive auxiliary hypothesis.

Results

The dependent variable was the relative preference for the carboxydismutase theory, calculated as the rating for the carboxydismutase theory minus the rating for the rubisco theory. There were two main effects with no interaction. First, there was a general preference for the rubisco theory, but this preference was less when the auxiliary hypothesis of the carboxydismutase theory was predictive rather than reactive (mean differences of -0.59 and -1.77, respectively; $F(1,72) = 4.39$, $p < .05$). Second, the carboxydismutase theory was relatively more believable with the revised-pathway auxiliary hypothesis than with the exception-barring hypothesis (mean differences of -0.29 and -2.11, respectively; $F(1,72) = 11.03$, $p < .01$).

These results support the predictivist thesis. The believability of the combination of the carboxydismutase theory and its auxiliary hypothesis (Theory Y + Auxiliary Hypothesis Y1) depended on when the auxiliary hypothesis was introduced. When the auxiliary hypothesis was presented as a reactive response to anomalous data, the carboxydismutase theory was less believable than when the auxiliary hypothesis was proposed prior to the discovery of the same data. The timing of theoretical hypotheses mattered, not just the theoretical hypotheses themselves.

Experiment 2

Experiment 2 provides a further test of the predictivist thesis, using a new domain and a new configuration of theories. The basic strategy was to contrast two situations, which are outlined in Table 2. In Situation A, Theory X and Theory Y are both designed to explain Data D, but the two theories make opposite predictions about a new piece of data, D1. Theory X predicts that D1 will be found, but Theory Y predicts that D1 will *not* be found. Theory X's prediction turns out to be correct, and Theory Y's is incorrect. In order to explain {D + D1}, proponents of Theory Y respond by adding auxiliary hypothesis Y1. In Situation B, Theory X and Theory {Y + Y1} are both developed after both D and D1 are already known; thus, both retrospectively explain the same data {D + D1}. In both situations, Theory X is contrasted, in the end, with Theory {Y + Y1} in the context of data {D + D1}. Thus, the final configuration of data and competing theories in the two situations is identical.

Table 2. Conceptual outline of Experiment 2.

Scenario	Situation A: Predictive	Situation B: Nonpredictive
time 1	Data D are known	Data D are known
time 2	Theory X is devised to explain D, and also predicts D1. Theory Y is devised to explain D, and also predicts *not* D1.	
time 3	Datum D1 becomes known.	Datum D1 becomes known.
time 4	Theory X needs no change. Theory Y adds hypothesis Y1 in order to explain D1.	Theory X is devised to explain D & D1. Theory Y with Y1 is devised to explain D & D1.
Final Configuration of Theories and Data	Theory X explains D & D1. Theory Y+Y1 explains D & D1.	Theory X explains D & D1. Theory Y + Y1 explains D & D1.

Hypotheses

According to the nonpredictivist position, subjects should not differ in their evaluations across the two situations. In both situations, Theory X is, in the end, pitted against Theory {Y + Y1}, and in both situations, X and {Y + Y1} are intended to explain the same data {D + D1}. Therefore, for the nonpredictivist, the history of X and Y (X's successful prediction and Y's reactive introduction of Y1) should not matter. Only the final configuration of X, {Y + Y1}, and {D + D1} should matter.

According to the predictivist view, the history of X and Y should matter. In Situation A, the fact that Theory X made a successful prediction and Y was put on the defensive should work to the favor of Theory X. Thus, Theory X should be relatively more attractive in Situation A than in Situation B.

Experiment 2 additionally raises the issue of whether different types of auxiliary hypotheses introduced in response to anomalous data can have different effects on theory evaluation. Perhaps predictivism holds for some types of auxiliary hypothesis but not for others.

Method

Subjects. Forty-four students in introductory educational psychology courses participated in the experiment.

Procedure and Materials. The basic procedure was like that of the previous experiment: Subjects read about two competing theories and then rated their belief in the two theories. In this experiment, the scientific issue was the cause of an extensive formation of deep channels that cover much of the eastern part of the state of Washington. The two competing theories were the single-flood theory, which holds that a single gigantic flood carved the channels, and the many-floods theory, which asserts that the channels were gradually cut by many small floods over thousands of years. The texts were 500 to 650 words long.

Design. Two factors were manipulated in a 2 x 2 design. The first factor was predictiveness. In the *predictive* condition, the single-flood theorists made a successful prediction, and the many-floods theorists made an unsuccessful prediction, which forced them to deal with the resulting anomalous data. In the *nonpredictive* condition, both theories were entirely retrospective, developed only after all of the relevant data were already known.

The second factor was the type of auxiliary hypothesis added by the many-floods theorists to account for the anomalous data. Two types of auxiliary hypotheses were used, a hypothesis that enabled the many-floods theorists to *reject* the data and a hypothesis that enabled the many-floods theorists to *accommodate* the data. Accommodating the anomalous data meant *accepting* the anomalous data as valid but *explaining the data away* without giving up the core hypothesis that many floods rather than a single flood carved the channels. In the reactive condition, the accommodative auxiliary hypothesis amounted to a *peripheral theory change* (see Chinn & Brewer, 1993; Lakatos, 1970). The resulting four conditions are described below:

1. Predictive/Reject. The single-flood theory makes a successful prediction; the many-floods theory makes an unsuccessful prediction. The many-floods theory rejects the anomalous data. According to the text in this condition, both theories were advanced in the 1920s to explain the channels in Washington state. The single-flood theory had very few proponents because single-flood theorists could not explain where all the water needed for a catastrophic flood would come from. The few single-flood proponents predicted that a source for the water would someday be discovered, but many-flood proponents were certain that no such source would ever be found. However, the single-flood theorists were vindicated when evidence was found for a large prehistoric lake in what is now western Montana. This lake had been dammed by a glacier, and the geologic evidence indicated that the lake had suddenly dumped enormous amounts of water to the west when the glacier dam broke apart. Thus, the single-flood theory had made a bold and successful prediction.

Many-floods theorists responded by rejecting this new piece of data. They argued that the flood waters never reached Washington because the waters were first swallowed up by fissures that opened in the ground under the lake. This response rejects the notion that the flood waters ever reached Washington.

2. Predictive/Accommodate. The single-flood theory makes a successful prediction; the many-floods theory makes an unsuccessful prediction. The many-floods theory accommodates the anomalous data, making an implicit change by denying that a single large flood could have carved the channels. The text in this experimental condition was identical to that of the previous condition, except for the response of the many-floods theorists to the evidence of the glacial flood. In this condition, unlike the previous condition, the many-flood theorists accepted the data indicating that the lake had flooded into Washington, so they needed to find a way to accommodate the data without giving up their theory. Therefore, they argued that even though a catastrophic flood had occurred, a single flood would not suffice to create deep channels, which can be created only by many years of repeated erosion.

This response appears to involve a modification to the many-floods theory. The fact that the many-flood theorists originally predicted that no source of water would be found implies that they had believed that a large flood could have carved the channels if there had been a large flood. Now this implied belief is altered, as the many-flood theorists now deny that a large flood could carve the channels.

3. Nonpredictive/Reject. Both theories are retrospective. The many-floods theory rejects the anomalous data. According to the text in this condition, both the single-flood theory and the many-floods theory were developed *after* the evidence for the glacial lake in Montana had been discovered. The single-flood theory cited the lake and its flooding to the west as strong supporting evidence. The many-floods theory rejected the data in the same way as in Condition 1, namely, by stating that fissures had opened up in the ground under the lake and had swallowed up the flood waters so that they never reached the Washington area.

4. Nonpredictive/Accommodate. Both theories are retrospective. The many-floods theory accommodates the anomalous data by denying that a single large flood could have carved the channels. As in Condition 3, both theories were said to have been developed *after* the evidence for the lake in Montana was discovered, and the single-flood theorists counted the lake as strong evidence for their theory. The many-flood theorists responded to the evidence for the lake as they did in Condition 2. They accepted that a large flood had occurred, but they argued that many repeated floods are necessary to produce deep channels, and that a single flood will not do the job. Notice, however, that unlike Condition 2, the denial that a single flood could cut the channels appears the first time that the subjects learn about the many-floods theory. There is no implied change to the theory.

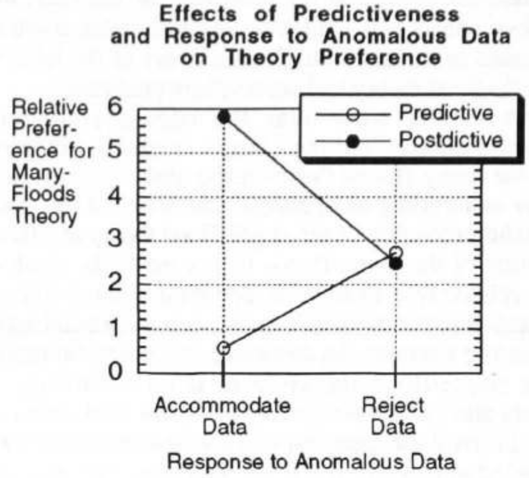

Figure 1. Results from Experiment 2

Results

The dependent variable was the relative preference for the many-floods theory, computed as the rating for the many-floods theory minus the rating for the single-flood theory. The results are shown in Figure 1. An ANOVA showed that the interaction was reliable ($F(1,40) = 8.25$, $p < .01$).

The interaction showed that the effects of the predictiveness of the two theories depended on whether the many-floods theorists rejected or accommodated the data. When the many-floods theorists rejected the anomalous data (Conditions 1 and 3), the predictiveness factor made no difference. Subjects preferred the many-floods theory by about 2.5 rating points regardless of whether the theories were predictive or nonpredictive.

However, when the many-floods theorists responded to the glacial lake data by denying that a single flood could cut the channels (Conditions 2 and 4), there was a very large difference in theory preference depending on whether the theories were predictive or nonpredictive. When (a) single-flood theorists successfully predicted a source of water and (b) many-floods theorists initially predicted that no source of water would be found and then reacted to the discovery of the lake by apparently altering their theory (now declaring that a single large flood could not cut deep channels), the single-flood theory was almost as believable as the many-floods theory. However, when the single-flood theory and the many-floods theory were both developed retrospectively after the glacial lake was already discovered (so that the many-floods theory declared from its inception that a single large flood could not cut deep channels), the many-floods theory was much more believable than the single-flood theory

Thus, although the final versions of the two rival theories were exactly the same in Conditions 2 and 4, subjects rated the theories very differently, depending on the histories of those theories. The auxiliary hypothesis that a single large flood cannot cut deep channels was quite believable when both theories postdated the discovery of the lake, but the same hypothesis was evidently less believable when it was formulated as a reaction to the discovery of the lake, which the single-flood theory had successfully predicted.

These results show that the timing of hypotheses sometimes matters, but not always. Predictivism prevailed when the many-floods theory responded to the glacial lake data by undergoing an apparent alteration. In this case, the successful prediction of the single-flood theory and the failed prediction of the many-floods theory made the single-flood theory relatively attractive. Nonpredictivism prevailed when the many-floods theorists responded to the glacial lake data by rejecting the data. In this case, the successful prediction of the single-flood theory gave it no advantage. This suggests that the advantages of successful predictions can be negated if rival theorists can find a way to reject the data that they failed to predict. But if the data cannot be rejected, the theory that makes a successful prediction gains an edge over rival theories that are forced to beat a retreat.

Conclusions

Many current cognitive models of theory choice have taken the nonpredictivist stance that theories are evaluated without reference to *when* the hypotheses within the theory were advanced. The predictivist thesis, on the other hand, insists that the timing of the introduction of hypotheses must be taken into account.

The experiments reported in this paper suggest that predictivism is at least sometimes correct. Successful predictions seem to give a theory an edge over a rival theory when the rival theory makes incompatible, and incorrect, predictions. Successful predictions seem to be advantageous under these general circumstances: Theory X and Theory Y make incompatible predictions. Theory X's prediction is correct, and Theory Y's prediction is incorrect, so Theory Y must be revised to Y* in order to accommodate the anomalous data. X gains an edge over Y* when Y* is formulated in response to X's successful prediction and Y's failed prediction; X does *not* gain an edge over the exact same Y* when Y* is formulated in a way that does not involve a failed prediction. Experiment 1 showed that in comparison with X, Y* is less believable when Y* is formulated in reaction to Y's failed prediction (and X's successful prediction) than when the same Y* is specified prior to the anomalous data. Experiment 2 showed that Y* is relatively less believable when Y* is formulated in reaction to Y's failed prediction (and X's successful prediction) than when both X and Y* are developed retrospectively. In both experiments, the evaluations of X and Y* depended on the histories of X and Y*. Thus, in at least some circumstances, it appears that cognitive models of theory choice must take the history of hypotheses into account.

In other circumstances, nonpredictivism was supported. This was true when Theory Y was not forced to retreat in the face of anomalous data. If Theory Y avoids the appearance of retreat by rejecting the anomalous data (Experiment 2) or by being developed *in toto* only after the data are already known (the pilot study), then Theory X's successful predictions do not appear to give it any advantage. Once again, it appears that successful predictions may be advantageous primarily when paired with unsuccessful predictions made by a rival theory. Future research should try to clarify the relative importance of Theory X's successful predictions and Theory Y's unsuccessful predictions in situations in which Theory X has a relative advantage.

Finally, it should be repeated that the experiments reported here used nonscientists as subjects. Further research is needed to determine whether these results would hold for scientists, as well.

Acknowledgments

I am grateful to Bill Brewer and Michael Ranney for their very helpful comments on an earlier version of this paper.

References

Brinberg, D., Lynch, J. G., Jr., & Sawyer, A. G. (1992). Hypothesized and confounded explanations in theory tests: A Bayesian analysis. *Journal of Consumer Research, 19,* 139-154.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1-49..

Horwich, P. (1982). *Probability and evidence.* Cambridge: Cambridge University Press.

Howson, C. (1991). The 'old evidence' problem. *British Journal for the Philosophy of Science, 42,* 547-555.

Howson, C., & Franklin, A. (1991). Maher, Mendeleev and Bayesianism. *Philosophy of Science, 58,* 574-585.

Keynes, J. M. (1921). *A treatise on probability.* London: Macmillan.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-196). London: Cambridge University Press.

Lakatos, I. (1976). *Proofs and Refutations.* Cambridge: Cambridge University Press.

Maher, P. (1990). Why scientists gather evidence. *British Journal for the Philosophy of Science, 41,* 103-119.

Maher, P. (1992). Diachronic rationality. *Philosophy of Science, 59,* 120-141.

Maher, P. (1993). Howson and Franklin on prediction. *Philosophy of Science, 60,* 329-340.

Popper, K. (1959). *The logic of scientific discovery.* London: Hutchinson.

Schank, P., & Ranney, M. (1992). Assessing explanatory coherence: A new method for integrating verbal data with models of on-line belief revision. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society,* pp. 599-604.

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