50 Years of Successful Predictive Modeling Should Be Enough: Lessons for Philosophy of Science

Michael A. Bishop and J. D. Trout†‡
Iowa State University and Loyola University, Chicago

Our aim in this paper is to bring the woefully neglected literature on predictive modeling to bear on some central questions in the philosophy of science. The lesson of this literature is straightforward: For a very wide range of prediction problems, statistical prediction rules (SPRs), often rules that are very easy to implement, make predictions than are as reliable as, and typically more reliable than, human experts. We will argue that the success of SPRs forces us to reconsider our views about what is involved in understanding, explanation, good reasoning, and about how we ought to do philosophy of science.

1. Introduction. Our aim in this paper is to bring the woefully neglected literature on predictive modeling to bear on some central questions in the philosophy of science. The lesson of this literature is straightforward: For a very wide range of prediction problems, statistical prediction rules (SPRs), often rules that are very easy to implement, make predictions than are as reliable as, and typically more reliable than, human experts. We will argue that the success of SPRs forces us to reconsider our views about what is involved in understanding, explanation, good reasoning, and about how we ought to do philosophy of science.

2. Statistical Prediction Rules. Prediction problems great and small are an

†Send requests for reprints to the authors. Bishop: Department of Philosophy and Religious Studies, 402 Catt Hall, Iowa State University, Ames, IA 50011; e-mail: mikebish@iastate.edu; Trout: Department of Philosophy, 6525 N. Sheridan Rd., Loyola University, Chicago, IL 60626; email: jtrout@orion.it.luc.edu.

‡We have received very valuable comments on earlier and partial drafts of this paper from Paul Abela, Joseph Mendola, Dominic Murphy, Jesse Prinz, Richard Samuels and the cognitive science group at Washington University, St. Louis.

Philosophy of Science, 69 (September 2002) pp. S197–S208. 0031-8248/2002/69supp-0018$10.00

essential part of everyday life. Most common prediction problems share a similar structure: On the basis of certain cues, we make judgments about some target property. Researchers have developed actuarial models (or SPRs) for various real-life prediction problems. These models provide a purely mechanical procedure for arriving at predictions on the basis of quantitatively coded cues. While there are many different kinds of SPRs, consider first *proper linear models* (Dawes 1982, 391). Suppose we want to predict the quality of the vintage for a red Bordeaux wine. A proper linear model for this prediction problem might take the following form:

$$P = w_1(c_1) + w_2(c_2) + w_3(c_3) + w_4(c_4)$$

where c_n is the value for the n^{th} cue, and w_n is the weight assigned to the n^{th} cue. For example, c_1 might reflect the age of the vintage, while c_2, c_3 and c_4 might reflect climatic features of the relevant Bordeaux region. To complete the proper linear model, we need a reasonably large set of data showing how these cues correlate with the target property (the market price of mature Bordeaux wines). Weights are then chosen so as to best fit the data: they optimize the relationship between P (the weighted sum of the cues) and the target property. An actuarial model along these lines has been developed (Ashenfelter, Ashmore, and Lalonde 1995). It predicts 83% of the variance in the price of mature Bordeaux red wines at auction. Reaction in the wine-tasting industry to such models has been "somewhere between violent and hysterical" (Passell 1990).

In his "disturbing little book" Paul Meehl (1954) asked the question: Are the predictions of human experts more reliable than the predictions of actuarial models? Meehl reported on 20 studies in which experts and actuarial models made their predictions on the basis of the same evidence (i.e., the same cues). Since 1954, almost every non-ambiguous study that has compared the reliability of clinical and actuarial predictions has supported Meehl's conclusion (Grove and Meehl 1996). So robust is this finding that we might call it *The Golden Rule of Predictive Modeling:* When based on the same evidence, the predictions of SPRs are at least as reliable, and are typically more reliable, than the predictions of human experts. SPRs have been proven more reliable than humans at predicting the success of electroshock therapy, criminal recidivism, psychosis and neurosis on the basis of MMPI profiles, academic performance, progressive brain dysfunction, the presence, location and cause of brain damage, and proneness to violence (for citations see Dawes, Faust, and Meehl 1989; Dawes 1994; Swets, Dawes, and Monahan 2000). Even when experts are given the results of the actuarial formulas, they still do not outperform SPRs (Leli and Filskov 1984; Goldberg 1968). Upon reviewing this evidence, Paul Meehl said:

> There is no controversy in social science which shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one. When you are pushing [scores of] investigations [140 in 1991], predicting everything from the outcomes of football games to the diagnosis of liver disease and when you can hardly come up with a half dozen studies showing even a weak tendency in favor of the clinician, it is time to draw a practical conclusion. (1986, 372–373)

Among the most important prediction problems we face are problems of *human and social* prediction. These prediction problems typically have the following features:

(1) Even the best SPRs are not especially reliable.
(2) The best cues are reasonably predictive and somewhat redundant.

When these conditions obtain, then the reliability of a linear model's predictions are not particularly sensitive to the weights assigned to the cues. This analytic finding in statistics is known as *the flat maximum principle* (Lovie and Lovie 1986). This principle has surprising implications. It implies that for human and social prediction problems, as long as you have the right cues, the reliability of your model is not particularly sensitive to what weights are assigned to the cues (except for the sign of the weights, of course). Among improper linear models, there is one that stands out for its ease of use and relative reliability. Unit weight models assign equal weights to standardized predictor cues, so that each cue has an equal "say" in the final prediction (Dawes and Corrigan 1974; Einhorn and Hogarth 1975).

3. SPRs: Success and Resistance. The sluggish reception SPRs have received in the disciplines whose business it is to predict and diagnose is puzzling. In the face of a half century of studies showing the superiority of SPRs, many experts still base judgments on subjective impressions and unmonitored evaluation of the evidence. Resistance to the SPR findings runs deep, and typically comes in the form of an instance of Peirce's Problem. Peirce (1878, 281–282) raised what is now the classic worry about frequentist interpretations of probability: How can a frequentist probability claim (say, that 99 out of 100 cards are red) be relevant to a judgment about a particular case (whether the next card will be red)? After all, the next card will be red or not, and the other 99 cards can't change that fact. Those who resist the SPR findings are typically quite willing to admit that *in the long run,* SPRs will be right more often than human experts. But their (over)confidence in subjective powers of reflection leads them to deny that we should believe the SPRs prediction *in some particular case.*[1]

1. Some complain that whenever experts and SPRs are compared, humans are forced

A legitimate worry about SPRs has come to be known as the "broken leg" problem. Suppose an actuarial formula accurately predicts an individual's weekly movie attendance. If we know that the subject has a broken leg, it would be wise to discard the actuarial formula (Dawes, Faust, and Meehl 1989). While broken leg problems will inevitably arise, it is difficult to offer any general prescriptions for how to deal with them. The reason is that in studies in which experts are given SPRs and are permitted to override them, the experts inevitably find more broken leg examples than there really are. In fact, such experts predict less reliably than they would have if they had just used the SPR (Goldberg 1968; Sawyer 1966; Leli and Filskov 1984). Our inclination is to suggest that overriding a SPR is a good idea only in very unusual circumstances. For example, there have been cases in which researchers came to realize that they could improve a SPR by adding more variables; in such cases, experts might well be able to improve upon the SPR's predictions by taking into account such evidence (Swets, Dawes, and Monahan 2000, 11).

In general, the resistance to the SPR findings are intimately bound up with our tendency to be overconfident about the power of our subjective reasoning faculties and about the reliability of our predictions. Our faith in the reliability of our subjective powers of reasoning bolsters our (over)confidence in our judgments; and our (over)confident judgments bolster our belief in the reliability in our subjective faculties. Let's focus on each side of this overconfidence feedback loop.

Overconfidence in our judgments. The overconfidence bias is one of the most robust findings in contemporary psychology.

> [A] large majority of the general public thinks that they are more intelligent, more fair-minded, less prejudiced, and more skilled behind the wheel of an automobile than the average person. . . . A survey of one million high school seniors found that 70% thought they were above average in leadership ability, and only 2% thought they were below average. In terms of ability to get along with others, *all* students thought they were above average, 60% thought they were in the top 10%, and 25% thought they were in the top 1%! Lest one think that such inflated self-assessments occur only in the minds of callow high-

---

to use only evidence that can be quantified. This allegedly rigs the competition in favor of the SPRs, because experts are not permitted to use the kinds of qualitative evidence that could prompt use of the experts' distinctly subjective human faculties. This complaint is bogus. It is perfectly possible to quantitatively code virtually any kind of evidence that is prima facie non-quantitative so that it can be utilized in SPRs. For example, the SPR that predicts the success of electroshock therapy employs a rating of the patient's insight into his or her condition. This is prima facie a subjective, non-quantitative variable because it relies on a clinician's diagnosis of a patient's mental state. Yet, clinicians can quantitatively code their diagnoses for use in a SPR.

> school students, it should be pointed out that a survey of university professors found that 94% thought they were better at their jobs than their average colleague. (Gilovich 1993, 77)

The overconfidence bias goes far beyond our inflated self-assessments. For example, Fischhoff, Slovic, and Lichtenstein (1977) asked subjects to indicate the most frequent cause of death in the U.S. and to estimate their confidence that their choice was correct (in terms of "odds"). When subjects set the odds of their answer's correctness at 100:1, they were correct only 73% of the time. Remarkably, even when they were so certain as to set the odds between 10,000:1 and 1,000,000:1, they were correct only between 85% and 90% of the time. It is important to note that the overconfidence effect is systematic (it is highly replicable and survives changes in task and setting) and directional (the effect is always in the direction of over rather than underconfidence).

What about scientists? Surely scientists' training and experience delivers them from the overconfidence bias in their areas of expertise. Alas, no—or at least, not always. Physicists, economists, and demographers have all been observed to suffer from the overconfidence bias, even when reasoning about the content of their special discipline (Henrion and Fischhoff 1986). It would appear that scientists place more faith in the subjective trappings of judgment than is warranted. Philosophers have supported this bad habit.

Overconfidence in the reliability of our subjective reasoning faculties. Humans are naturally disposed to exaggerate the powers of our subjective faculties. A very prominent example of this is the interview effect. When gatekeepers (e.g., hiring and admissions officers) are allowed personal access to applicants in the form of unstructured interviews, they are still outperformed by SPRs that take no account of the interviews. In fact, unstructured interviews actually degrade the reliability of human prediction (for citations, see Dawes 1994). That is, gatekeepers degrade the reliability of their predictions by availing themselves of unstructured interviews.

Although the interview effect is one of the most robust findings in psychology, highly educated people ignore its obvious practical implication. This occurs because of Peirce's Problem and our confidence in our subjective ability to "read" people. We suppose that our insight into human nature is so powerful that we can plumb the depths of a human being in a 45 minute interview—unlike the lesser lights who were hoodwinked in the SPR studies. Our (over)confidence survives because we typically don't get systematic feedback about the quality of our judgments (e.g., we can't compare the long-term outcomes of our actual decisions against the decisions we would have made if we hadn't interviewed the candidates). To put this in practical terms, the process by which most contemporary philosophers were hired was seriously and, at the time, demonstrably flawed. This will be of no comfort to our colleagues, employed or unemployed. We expect, however, that the unemployed will find it considerably less surprising.

We do not intend to offer a blanket condemnation of the overconfident. We recognize that overconfidence may be a trait that is essential to psychic health. It may be one of nature's ways of helping us cope with life's inevitable setbacks (Taylor 1989). As such, overconfidence may also sometimes play a useful role in science, e.g., it might lead a young turk to defend a promising new idea against the harsh objections of a well developed tradition. We have harped on our overconfidence so that we may preempt certain kinds of opposition—or at least try to. In the following sections, we will object to the epistemological role that subjective, internalist notions have played in various philosophical theories. While there may be many legitimate objections to what we have to say, it is surely uncontroversial that an unjustified, resolute overconfidence in the reliability of our subjective reasoning faculties is an appalling foundation on which to base any serious philosophical theory or method.

4. The Nature of Explanation. The epistemology of explanation is a two-headed monster. Most of the widely discussed accounts of explanation have been objectivist: What makes an explanation good concerns a property that it has independent of explainers; it concerns features of external objects, independent of particular minds. At the same time, virtually all contemporary accounts of explanation agree on one point: Understanding is centrally involved in explanation, whether as an intellectual goal or as a means of unifying practice. As philosophers of explanation are not chiefly in the business of analyzing traditional epistemic concepts, their notions of understanding and justification reflect a default internalism. This ordinary internalism includes an internal access condition that justification determiners must be *accessible to,* or *knowable by,* the epistemic agent. This internal accessibility is thought to contribute to, if not constitute, the agent's understanding. Accordingly, this unvarnished internalism implies that it is a necessary condition for us to be justified that we understand the contents that we are representing. Only then can we act on those contents responsibly. The conception of justification that is grounded in understanding isolates reason-giving as the characteristic model of justification—justification as argument.

It is in terms of this default internalism, then, that we should interpret claims about understanding expressed by philosophers of science. Peter Achinstein (1983, 16) asserts a "fundamental relationship between explanation and understanding." Wesley Salmon (1998, 77) proposes that scientific understanding is achieved in two ways: by "fitting phenomena into a comprehensive scientific world-picture," and by detailing and thereby exposing the "inner mechanisms" of a process. Michael Friedman (1974, 189) claims that the relation of phenomena that "gives understanding of the explained phenomenon" is "the central problem of scientific explanation." Philip Kitcher (1981, 168) relates understanding and explanation so closely that elucidation of this connection in a theory of explanation "should show us *how* scientific explanation advances our understanding." James Woodward (1984, 249) claims that a theory of explanation should "identify the structural features of such explanation which function so as to produce understanding in the ordinary user." None of these accounts, however, have much to say about the precise nature of understanding. Perhaps these positions rest the centrality of understanding on the consensus that there *is* such a thing as understanding. But the cognitive relation or state of understanding is itself a proper object of scientific inquiry, and its study—or the study of the components that comprise it—is actually carried out by cognitive psychology (Trout 2002).

But if explanatory scientific understanding requires seeing "how we can fit them [phenomena] into the general scheme of things, that is, into the *scientific world-picture*" (Salmon 1998, 87), then most people are incapable of explanatory scientific understanding, including most scientists. Indeed, when scientists piece together phenomena, they do so by focussing on the detailed findings of their (usually) narrow specialization. In contemporary science, global unification arises spontaneously from coordinated piecemeal efforts, not from a meta-level at which the philosopher or reflective scientist assembles remote domains (Miller 1987). Indeed, in light of the arcaneness of contemporary theoretical knowledge, no single individual can be so situated. Accordingly, actual explanatory practices in science appear to violate the internal access condition, and thus must be far more externalist than current accounts of explanation suppose.

It is not just *philosophical* theories of explanation that have accorded to the sense of understanding an essential role in explanation. Psychological theories of explanation, too, appeal to a sense of understanding, in both everyday and scientific explanation. Like some global, unifying accounts of explanation in the philosophy of science, a prominent psychological account focuses on the unified conceptual framework it provides: "in everyday use an explanation is an account that provides a conceptual framework for a phenomenon (e.g., fact, law, theory) that leads to a feeling of understanding in the reader-hearer" (Brewer et al. 1998, 120). And scientific explanations are no different in this respect; they should "provide a feeling of understanding" (121).

These psychological descriptions of understanding focus on its phenomenology. There is "something that it is like" to understand, and we use the precise character of this subjective sense that we understand—a psychological impression of coherence, confidence, etc.—as a cue that we do indeed understand. But the sense of understanding no more means that you have knowledge of the world than caressing your own shoulder means that someone loves you. Just ask Freud.

5. Methodology in the Philosophy of Science. Contemporary philosophers and historians of science who propose general hypotheses about how science works typically rely on case studies. They recruit episodes from the history of science that are confirming instances of their hypotheses. However naturalistic, this approach to the philosophy of science is relentlessly narrative. The point is to tell stories about episodes in the history of science that instantiate some principle (e.g., a methodological principle like "parsimony is a crucial factor in theory-choice"). These compelling narratives might well give us a subjective sense that we have grasped some deep truth about the nature of how science operates. But as we have argued, it is a serious mistake to suppose that such trappings of subjective judgment are a reliable sign of genuine understanding. Further, the hypothesis about how science works might fit coherently with all the relevant evidence known. But again, responsible reasoning need not attend closely to the satisfaction of such internalist virtues.

How much support does a single case study (or even a number of case studies) provide a general principle about the nature of science? This question cannot be answered with armchair speculation, no matter how scrupulous. When faced with a case study that supports some hypothesis, we need to know the *relative* frequency of such supporting cases (compared to those that might disconfirm it). After all, for any general hypothesis about the nature of science some professional philosopher or historian has defended, it is possible that there is some episode in the history of science that confirms it and some other that disconfirms it. We also need to know base-rate information about the occurrence of such episodes (Trout 1998). How prevalent is the phenomenon described by the general principle?

It would be a monumental task to try to figure out the relative frequency or the base rate of some phenomenon in the history of science. Indeed, one is not sure how to even begin: How do we individuate episodes? What features do we consider in coding them? How do we select which ones to consider? These are difficult questions that must at least be addressed before the necessary quantitative, actuarial work gets done (Faust and Meehl 1992). But here is an interesting fact that might give us pause: On at least one way of counting, about 90% of all scientists who have ever lived are alive today. It is jarring to note that the vast majority of published case studies describe the activities of the 10% of dead scientists. Needless to say, it is dangerous to extract relative frequency or base-rate conclusions from such a non-random sample. And yet one worries that those experts with the greatest knowledge of published case studies, and whose judgments are likely to be most confident and receive the most deference, are doing just that.

An actuarial approach to the history and philosophy of science draws upon, and is subject to evaluation by, the best science of the day; it is therefore squarely naturalistic. It is ironic that naturalistic philosophers—philosophers who are inclined to see no principled methodological gaps separating science and philosophy—employ a method for confirming generalizations that, from a scientific perspective, is highly unsophisticated. (For two egregious, and indeed scandalous, examples of the improper use of case studies, see Bishop 1999 and Trout 1994.) Of course, given the daunting issues that must be addressed and resolved before we even begin to approach the philosophy of science from an actuarial perspective, it is perhaps understandable that we philosophers have avoided careful scrutiny of our case study methods. But perhaps the time has arrived for us either to give up the traditional narrative case study method in favor of an actuarial approach, or to explain how the traditional method is consistent with our avowals of respect for the empirical findings and methodological dictates of our best science.

6. SPRs and Epistemology. Epistemology in the twentieth century has been dominated by internalism. Internalism holds that what determines the justificatory status of a belief is in some sense internal to, or in principle knowable by, a believer. There is active disagreement among internalists about the precise nature of the internal, epistemic virtue that fixes the justificatory status of a belief. But live candidates include coherence, having good reasons, and fitting or being supported by the evidence.

We contend that the SPR results cast serious doubt on any kind of epistemic internalism that claims to be action-guiding. A prescriptive internalism will presumably advise reasoners to (other things being equal) adopt (or try to adopt) beliefs that best exhibit the internalist's favored epistemic virtue(s). But consider a case in which S, who knows the SPR results, is faced with a social prediction problem but does not have the wherewithal to construct or implement a proper linear model. S has a choice to use an improper unit weight model or to reason the way human experts often do—by considering a number of different lines of evidence and trying to weigh these lines of evidence in accordance with their predictive power. We contend that for many such problems, an action-guiding internalism yields the result that the epistemically responsible reasoner must use her subjective reasoning powers rather than rely on an improper unit weight model.

But surely any epistemic view is mistaken that recommends that we use our subjective reasoning powers rather than the simpler, more reliable unit weight model. Given a choice between two reasoning strategies where one of them is known to be less reliable *and* harder to use, to insist that the responsible reasoner adopt that strategy is to insist upon epistemic masochism, not epistemic responsibility.

How shall we reduce the effects of bias so rampant in traditional epistemology? There are two proposals. An *inside strategy* for debiasing attempts to improve the accuracy of judgment by creating a fertile corrective environment *in the mind.* A behavioral policy based on an inside strategy permits the alcoholic to sit at the bar and rehearse the reasons to abstain. An *outside strategy* identifies a principle or rule of conduct that produces the most accurate or desirable available outcome, and sticks to that rule despite the subjective pull to abandon the principle. A behavioral policy based on an outside strategy recommends that you avoid the bar in the first place. This outside, "policy" approach to decision-making might require that you select a solution that is not intuitively satisfying, but is objectively correct (Kahneman and Lovallo 1993).

Does the decisive success of this class of outside strategies—binding ourselves to the use of SPRs—imply either that subjective judgment is always unreliable, or that theoretically untutored notions are always scientifically disreputable? No. But the success of SPRs certainly doesn't help the case for internalism either. Outcome information is the chief, if not the sole, determinant of whether a method can be accurately applied. The feeling that we understand, the confidence that we have considered all of the relevant evidence, the effort and concentration on theoretical detail—in short, all of the subjective trappings of judgment—these are now known to be inferior predictors of accuracy than SPRs in the fields discussed. In some historical moments, ideologues have opined that a method or instrument that was in fact more accurate than those extant were less preferable for narrowly religious reasons concerning a local doctrine, or for narrowly political reasons concerning oppressive norms. But these arguments are difficult to sustain while endorsing methodological rigor and the in-principle defeasibility of any empirical claim, ideological or not. For those who are contemptuous of science, perhaps there is no cure. But for the rest of us, it is time to take our medicine.

This focus on outcomes means that, without relying on outcome information in such domains as psychotherapy, oncology, the psychology of criminal behavior, etc., "expert" claims originating in subjective evaluation can be safely ignored for what they are: sentimental autobiography. We cannot begin to repair the damage done by our indulgence of these internalist conceits, conceits that have persisted beyond the decades that exposed them. Incorrectly diagnosed cancers, dangerous criminals released, innocent people put to death, needless neglect of progressive brain disease, the misidentification of psychotics—and the wine, my God the wine—these failures demand frank admission. Anyone for absolution?

REFERENCES

Achinstein, Peter (1983), *The Nature of Explanation.* New York: Oxford University Press.

Ashenfelter, Orley, David Ashmore, and Robert Lalonde (1995), "Bordeaux Wine Vintage Quality and the Weather", *Chance* 8: 7–14.

Bishop, Michael (1999), "Semantic Flexibility in Scientific Practice: A Study of Newton's Optics", *Philosophy and Rhetoric* 32: 210–232.

Brewer, William, Clark Chinn, and Ala Samarapungavan (1998), "Explanation in Scientists and Children", *Minds and Machines* 8: 119–136.

Dawes, Robyn (1982), "The Robust Beauty of Improper Linear Models in Decision-Making" in Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press, 391–407.

——— (1994), *House of Cards: Psychology and Psychotherapy Built on Myth.* New York: The Free Press.

Dawes, Robyn, and Bernard Corrigan (1974), "Linear Models in Decision Making", *Psychological Bulletin* 81: 95–106.

Dawes, Robyn, David Faust, and Paul Meehl (1989), "Clinical Versus Actuarial Judgment", *Science* 243: 1668–1674.

Einhorn, Hillel J. and Robin M. Hogarth (1975), "Unit Weighting Schemas for Decision Making", *Organizational Behavior and Human Performance* 13: 172–192.

Faust, David and Paul Meehl (1992), "Using Scientific Methods to Resolve Enduring Questions within the History and Philosophy of Science: Some Illustrations", *Behavior Therapy* 23: 195–211.

Faust, David and Jay Ziskin (1988), "The Expert Witness in Psychology and Psychiatry", *Science* 241: 1143–1144.

Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein (1977), "Knowing with Certainty: The Appropriateness of Extreme Confidence", *Journal of Experimental Psychology: Human Perception and Performance* 3: 552–564.

Friedman, Michael ([1974] 1988), "Explanation and Scientific Understanding." Reprinted in J. C. Pitt (ed.), *Theories of Explanation.* New York: Oxford University Press, 188–198. Originally published in *Journal of Philosophy* 71: 5–19.

Gilovich, Thomas (1991), *How We Know What Isn't So.* New York: The Free Press.

Goldberg, Lewis (1968), "Simple Models of Simple Processes? Some Research on Clinical Judgments", *American Psychologist* 23: 483–496.

Grove, William M. and Paul E. Meehl (1996), "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy", *Psychology, Public Policy, and Law* 2: 293–323.

Henrion, Max and Baruch Fischhoff (1986), "Assessing Uncertainty in Physical Constants", *American Journal of Physics* 54: 791–798.

Kahneman, Daniel and Dan Lovallo (1993), "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking", *Management Science* 39: 17–31.

Kitcher, Philip ([1981] 1988), "Explanatory Unification." Reprinted in J. C. Pitt (ed.), *Theories of Explanation.* New York: Oxford University Press, 167–187. Originally published in *Philosophy of Science* 48: 507–531.

Leli, Dano A. and Susan B. Filskov (1984), "Clinical Detection of Intellectual Deterioration Associated with Brain Damage", *Journal of Clinical Psychology* 40: 1435–1441.

Lovie, Alexander D. and Patricia Lovie (1986), "The Flat Maximum Effect and Linear Scoring Models for Prediction", *Journal of Forecasting* 5: 159–168.

Meehl, Paul (1954), *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* Minneapolis: University of Minnesota Press.

———(1986), "Causes and Effects of My Disturbing Little Book", *Journal of Personality Assessment* 50: 370–375.

Miller, Richard W. (1987), *Fact and Method.* Princeton: Princeton University Press.

Passell, Peter (1990), "Wine Equation Puts Some Noses Out of Joint", *The New York Times,* March 4, p. 1.

Peirce, Charles S. ([1878] 1982), "Doctrine of Chances", in *Writings of Charles Sanders Peirce: A Chronological Edition,* vol. 3. Bloomington: Indiana University Press, 276–289.

Salmon, Wesley (1998), "The Importance of Scientific Understanding", in *Causality and Explanation.* New York: Oxford University Press, 79–91.

Swets, John A., Robyn Dawes, and John Monahan (2000), "Psychological Science Can Improve Diagnostic Decisions", *Psychological Science in the Public Interest* 1: 1–26.

Taylor, Shelly (1989), *Positive Illusions: Creative Self-Deception and the Healthy Mind.* New York: Basic Books.

Trout, J. D. (1994), "A Realistic Look Backward", *Studies in History and Philosophy of Science* 25: 37–64.

——— (1998), *Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences.* New York: Oxford University Press.

——— (2002), "Scientific Explanation and the Sense of Understanding", *Philosophy of Science* 69: 212–234.

Tversky, Amos and Daniel Kahneman (1986), "Rational Choice and the Framing of Decisions", *Journal of Business* 59: S251–S278.

Woodward, James. ([1984] 1993), "A Theory of Singular Causal Explanation." Reprinted in D. H. Ruben (ed.), *Explanation.* New York: Oxford University Press, 246–274. Originally published in *Erkenntnis* 21: 231–262.