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

J.D. Trout & Michael Bishop

DRAFT: Comments are welcome & may be sent to mikebish@iastate.edu

The aims of this paper are two, though the second aim has at least 4 parts. Our first aim is to briefly introduce the fascinating and important literature on predictive modelling (section 1). The lesson of this literature – of over a half century's worth of studies – is simple and straightforward: For a very wide range of prediction problems, statistical prediction rules (or SPRs), often rules that are very easy to implement, make more reliable predictions than human experts. This literature has been mistakenly rejected by many otherwise reasonable folk (section 2), and it has been woefully neglected by contemporary ethicists, epistemologists and philosophers of science (sections 3−6). Our second aim is to try to make up for lost time by bringing this literature to bear on some central philosophical questions. We argue that the success of SPRs forces us to reject the internalist accounts of justification and good reasoning that currently dominate epistemology (section 3) and replace them with an artless commitment to accuracy and reliability (sections 3 and 6). The success of SPRs also forces us to reconsider the role of understanding in philosophical accounts of explanation (section 4), and it casts serious doubt on the relentlessly narrative case study method philosophers and historians of science often use to assess general hypotheses about the nature of science (section 5). If the SPR results bring in their wake even a fraction of these implications, then we can expect revolutionary changes in our views about what's involved in understanding, explanation and good reasoning, and therefore in our views about how we ought to do philosophy of science.

1. Statistical Prediction Rules (SPRs)

Prediction problems great and small are an essential part of everyday life. What menu items will I most enjoy eating? Is this article worth reading? Is the boss in a good mood? Will the bungee cord snap? These and other common prediction problems share a similar structure: On the basis of certain cues, we make judgments about some target property. I doubt the integrity of the bungee cord (target property) on the basis of the fact that it looks frayed and the assistants look disheveled and hungover (cues). How we make such evidence−based judgments, and how we ought to make them, are interesting issues in their own right. But these issues are particularly pressing because such predictions often play a central role in decisions and actions. Because I don't trust the cord, I don't bungee jump off the bridge.

Researchers have developed many actuarial models for various real−life prediction problems. These actuarial models provide a purely mechanical procedure for arriving at a prediction on the basis of quantitatively coded cues. While there are many different kinds of actuarial models, we will focus first on proper linear models (Dawes 1979/82, 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 (the warmth of the growing season, the precipitation in August and September, and the previous winter's precipitation). 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. As the reader might have guessed, 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 1954, Paul Meehl wrote a classic book entitled, Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Literature. Meehl asked a simple question: Are the predictions of human experts more reliable than the predictions of actuarial models? To be a fair comparison, both the experts and the models had to make their predictions on the basis of the same evidence (i.e., the same cues). Meehl reported on 20 such experiments. Since 1954, every non−ambiguous study that has compared the reliability of clinical and actuarial predictions (i.e., Statistical Prediction Rules, or SPRs) has supported Meehl's conclusion. 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 more reliable than the predictions of human experts.

It is our contention that The Golden Rule of Predictive Modeling has been woefully neglected. Perhaps a good way to begin to undo this state of affairs is to briefly describe ten of its instances. This will give the reader some idea of the range and robustness of the Golden Rule.

1. A SPR that takes into account a patient's marital status, length of psychotic distress, and a rating of the patient's insight into his or her condition predicted the success of electroshock therapy more reliably than a hospital's medical and psychological staff members (Wittman 1941).

2. A model that used past criminal and prison records was more reliable than expert criminologists in predicting criminal recidivism (Carroll 1982).

3. On the basis of a Minnesota Multiphasic Personality Inventory (MMPI) profile, clinical psychologists were less reliable than a SPR in diagnosing patients as either neurotic or psychotic. When psychologists were given the SPR's results before they made their predictions, they were still less accurate than the SPR (Goldberg 1968).

4. A number of SPRs predict academic performance (measured by graduation rates and GPA at graduation) better than admissions officers. This is true even when the

admissions officers are allowed to use considerably more evidence than the models (DeVaul et al. 1957), and it has been shown to be true at selective colleges, medical schools (DeVaul et al. 1957), law schools (Dawes, Swets and Monohan 2000, 18) and graduate school in psychology (Dawes 1971).

5. SPRs predict loan and credit risk better than bank officers. SPRs are now standardly used by banks when they make loans and by credit card companies when they approve and set credit limits for new customers (Stillwell et. al. 1983).

6. SPRs predict newborns at risk for Sudden Infant Death Syndrome (SIDS) much better than human experts (Lowry 1975; Carpenter et. al. 1977; Golding et. al. 1985).

7. Predicting the quality of the vintage for a red Bordeaux wine decades in advance is done more reliably by a SPR than by expert wine tasters, who swirl, smell and taste the young wine (Ashenfelter, Ashmore and Lalonde 1995).

8. A SPR correctly diagnosed 83% of progressive brain dysfunction on the basis of cues from intellectual tests. Groups of clinicians working from the same data did no better than 63%. When clinicians were given the results of the actuarial formula, clinicians still did worse than the model, scoring no better than 75% (Leli and Filskov 1984).

9. In predicting the presence, location and cause of brain damage, a SPR outperformed experienced clinicians and a nationally prominent neuropsychologist (Wedding 1983).

10. In legal settings, forensic psychologists often make predictions of violence. One will be more reliable than forensic psychologists simply by predicting that people will not be violent. Further, SPRs are more reliable than forensic psychologists in predicting the relative likelihood of violence, i.e., who is more prone to violence (Faust and Ziskin 1988).

Upon reviewing this evidence in 1986, 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" (Meehl 1986, 372−3).

Among the most important prediction problems we face are problems of <u>human and social</u> prediction. Which applicant will be the best teacher, student, salesperson? Will this applicant repay this loan? If this prisoner is paroled, will he commit a violent crime? Problems of human and social prediction typically have the following features:

(1) Even the best SPRs are not especially reliable.
(2) The best cues are reasonably predictive.
(3) The cues are somewhat redundant (e.g., the larger a loan seeker's salary, the less likely she is to have claimed bankruptcy).

When these conditions obtain (no matter what the subject matter of the prediction problem), 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 prediction problems that satisfy conditions 1−3, 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). To see just how counterintuitive this implication is, consider three kinds of improper linear models.

Bootstrapping models. Goldberg (1970) gave 29 psychologists a series of MMPI profiles and asked them to predict whether patients would be diagnosed as neurotic or psychotic. Then for each psychologist, he constructed a bootstrapping model – a proper linear model that mimics the psychologist's predictions. In other words, he constructed 29 proper models that would take as cues the MMPI profile scores and as the target property a psychologist's predictions. Then Goldberg tested the bootstrapping models against the psychologists they aped. One might expect that bootstrapping models would predict nearly as well as the human expert on which they are based. But Goldberg found that in 26 of the 29 cases, the bootstraping model was more reliable in its diagnoses than the psychologist on which it was based! In other words, the bootstrapping model is built to ape an expert's predictions. But when it's wrong about the expert, it's more likely than the expert to be right about the target property.

Random linear models. Dawes and Corrigan (1974) took five successful bootstrapping models. For each model, they replaced each weight with a randomly chosen weight with the same sign. (So if the original model takes a cue to be positively [negatively] correlated with the target property, the random model would also reflect that correlation.) The random models were about as reliable as the bootstrapping models and more reliable than humans.

Unit weight models. Among improper linear models, there is one that tends to stand 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, Lovie and Lovie 1986). For problems of human and social prediction, unit weight models are about as reliable as proper models, and more reliable than expert humans.

2. SPRs: Success and Resistance

Proper models are very reliable because (a) the variables in proper models are correlated with the target property, (b) the values of those variables accurately reflect the real values of objects, and (c) the variables are weighted so as to best fit a large set of data. But why are improper (bootstrapping, random, unit weight) models so reliable? The answer is that in most practical situations, as long as (a) and (b) obtain, (c) doesn't have to. That's the lesson of the flat maximum principle. For many prediction problems of practical importance, as long as your linear model is looking at the right cues, and

your weights have the right (positive or negative) signs next to them, the reliability of the model won't be much affected by what weights you choose.

It is difficult to overstate just how powerful these results are, though researchers have done their best. For example, Paul Meehl has said that "[i]n most practical situations an unweighted sum of a small number of 'big' variables will, on the average, be preferable to regression equations" (quoted in Dawes and Corrigan 1974, 105). Dawes and Corrigan say that to be more reliable than expert humans in the social arena, "the whole trick is to know what variables to look at and then know how to add" (1974, 105). To put this yet another way: If the admissions officers of your college or university do not use SPRs, you can admit stronger students from a pool of applicants (students who will have relatively higher graduation rates and GPA's) simply by adding up each applicant's high school rank (out of 100) and their aptitude test score rank (out of 100) and admitting the students with the highest totals.[1,2]

The sluggish reception SPRs have received in the disciplines whose business it is to predict and diagnose is puzzling.[3] In the face of a half century of experiments 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 very deep, and typically comes in the form of an instance of Pierce's Problem. Pierce (1878, 281−2) raised what is now the classic worry about frequentist interpretations of probability: How can a probability claim (say, the claim 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 <u>in the long run</u>, 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 SPR's prediction <u>in some particular case</u>. Robyn Dawes recounts numerous cases in which people resist SPRs. For example, Dawes implemented a simple actuarial formula for predicting psychosis or neurosis on the basis of an MMPI profile at the Ann Arbor VA Hospital. "The single most effective rule for distinguishing the two conditions was quite simple: add scores from three scales and then subtract scores from two other scales. If the sum falls below 45, the patient is diagnosed as neurotic; if it equals or exceeds 45, the patient is diagnosed psychotic. This has come to be known as the 'Goldberg Rule'" (Dawes, Faust and Meehl 1989, 1669). Dawes describes clinicians's reaction to the formula.

> Whenever the clinicians in the hospital found a patient who had clearly been misclassified by this formula, they pointed that error out to me, sometimes gleefully… They were silent about the errors they made that the formula didn't; perhaps they did not even note them. The result was that their memory was biased against the formula and in their own favor. I was confidently assured that the formula didn't work as well as I had maintained… as if the clinicians' memory of a small sample of patients were a better basis for establishing the formula's validity than a sample of more than a thousand patients analyzed systematically. (When I pointed out this possible bias in their evaluation, my colleagues would good−

> naturedly agree that it presented a problem, but none were motivated to do a systematic study of the accuracy of their own judgment, even on the small sample available.) (Dawes 1994, 85−6)

Dawes recounts another vivid example. He was presenting a finding in which a SPR had outperformed various medical doctors in predicting the severity of disease and death. In the question period, "the dean of a prestigious medical school stated during the question period that 'if you had studied Dr. So−and−so, you would have found that his judgments of severity of the disease process would have predicted the survival time of his patients.' I could not say so, either publicly or privately, but I knew that the physician involved in fact was Dr. So−and−so…" (Dawes 2000, 151).

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 bolsters 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−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. Further, philosophers have supported this habit. Many epistemologists defend views of justification that favor subjective notions of coherence, support, and fit with evidence over brute reliability. Philosophers of science are guilty twice over. First, many defend views of understanding and explanation that give pride of place to the machinery of subjective judgment. And second, in drawing lessons about how science works, philosophers and historians of science often employ a relentlessly subjective, narrative approach. This approach relies on our subjective sense of having understood a particular historical episode and on generalizing that understanding to other cases (without benefit of any base−rate information or any information about the representativeness of the episode).

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 (Bloom and Brundage 1947; DeVaul et al. 1957; Oskamp 1965; Milstein et al. 1981). 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 experiments. 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 want 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 four sections, we will object to the epistemological role that subjective, internalist notions have played in philosophical accounts of good reasoning (section 3) and of explanation (section 4); we

will object to philosophers' reliance on relentlessly subjective, narrative methods in defending generalizations about the nature of science (section 5); and we will consider the ethical implications of relying on "feel good" subjective prediction rather than actuarial prediction for decisions of significant practical or social import (section 6). 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 build any serious philosophical theory.

3. Responsible reasoning

Suppose someone has some choice about what reasoning strategy to adopt in tackling a problem. Ignoring normative but non−epistemic (i.e., moral and pragmatic) considerations, how ought she to reason?[4] This epistemic "ought" is intended to be essentially prescriptive. It is useful and intuitive to suppose that this prescriptive function can be carried out by our notion of epistemic responsibility.[5]

1. Ignoring normative but non−epistemic (i.e., moral and pragmatic) considerations, when faced with a reasoning problem, one ought to reason in the most epistemically responsible manner.

We will argue that the SPR findings imply a kind of reliabilism about epistemic responsibility. While reliabilism is a well−known view about epistemic justification (Goldman 1986), we suggest that it is a better view about epistemic responsibility (see Bishop, in progress).

Responsibility reliabilism assesses voluntary reasoning strategies in terms of the use to which a reasoner is likely to put the strategy. Once we know the kinds of problems S is likely to try to solve using the mechanism, we can (in principle, at least) test its reliability on a large random sample of such problems. So suppose there are *m* psychologically real characterizations of the voluntary belief−forming strategy S uses to solve an empirical reasoning problem (where *m* might be 1). Each of these will define a process that has a reliability score, r_m. How responsible it is for S to use a belief−forming process, *p*, is a function of r_p, its reliability score. Now take the psychologically plausible mechanism (or mechanisms) with the highest reliability rating for that sample of problems. Psychologically plausible here does not just mean psychologically possible. Any mechanism is plausible that requires no greater resources than it would be reasonable for the subject to devote to this problem (reasonable on non−epistemic grounds, e.g., moral and instrumentally rational grounds). The most reliable, plausible belief−forming mechanism sets the standard of ideal epistemic responsibility. A subject's reasoning is more or less responsible to the extent that her mechanism's reliability rating departs from the ideal.

Perhaps the biggest advantage of reliabilism about responsibility rather than reliabilism about justification is that the former view avoids the generality problem. The generality problem arises because there are many ways to characterize a belief−forming mechanism. Some characterizations will denote a reliable process; others won't. This is

a problem for reliabilism about justification because it is a theory for assessing *belief tokens*: the justificatory status of a belief is a function of the reliability of the process that produced it. Thus, the theory requires a unique characterization of that process – otherwise the reliabilist will sometimes be stuck saying that a belief is both justified and unjustified. And that's absurd (Goldman 1979, Feldman 1985). But reliabilism about epistemic responsibility is a theory for assessing an *event* – an episode of reasoning (or, perhaps better, the implementation of a reasoning strategy). Different episodes of reasoning can have different, incompatible epistemic properties. So there is no need for the reliabilist about responsibility to demand a unique characterization of the process that produces a belief token.

Epistemic responsibility, as characterized above, is interesting for the same reason that epistemology is interesting: it tells us how we ought to reason. Epistemology is not merely an abstract, theoretical endeavor. Different views about how we ought to reason might well recommend different reasoning strategies for those charged with making decisions of lasting practical importance, including parole boards, AIDS diagnosticians, bank loan officers, hiring officers, university admissions committees, etc. Let's consider a rather prosaic social prediction problem. Hobart and Lance are admissions officers who are perfectly well acquainted with the flat maximum principle and its implications. They are trying to decide on the basis of college applications whether Smith or Jones will be the stronger student. Like the vast majority of reasoners, they do not have the wherewithal to construct or implement a proper linear model for this problem. Hobart employs a unit weight model, in which only two lines of evidence (high school rank and aptitude test score rank) are considered. Lance considers Hobart's two lines of evidence, as well as other lines of evidence (e.g., high school transcripts, letters of recommendation, extracurricular activities), and does his best to weigh these lines of evidence in accordance with their predictive power. Who is being more epistemically responsible? It seems clear that the reliabilist view of responsibility sketched above gives us the right answer: Hobart is the responsible one. Epistemic responsibility is essentially action−guiding, and from an epistemological perspective, one ought to employ Hobart's unit weight model. It is, after all, the reasoning strategy that both Lance and Hobart know is more reliable and easier to use. To argue that one ought to adopt Lance's reasoning strategy instead, when it is less reliable and harder to use, is to insist upon epistemic masochism, not epistemic responsibility.

2. In this example, Hobart's predictions are epistemically responsible, and Lance's predictions are epistemically irresponsible.

Now let's turn to what we will call internalist epistemic virtues. These are epistemic virtues that internalists take to be central to epistemic justification. Internalists believe that what determines the justificatory status of a belief is in some sense internal to, or in principle knowable by, a believer. Internalist virtues include coherence, having good reasons, and fitting the evidence. While there are interesting and important differences between these (and other) internalist virtues, for our purposes, we can associate such virtues with the predictions of proper linear models. Recall that the proper linear model's predictions are the result of considering all the different lines of relevant, available evidence and then weighing each line of evidence according to its predictive

value. A prediction made by the proper model optimizes a belief−system's coherence, it best fits the available evidence, it has the best reasons in its favor, etc.[6]

3. For the above prediction problem (and many others), the prediction of a well−constructed proper linear model best satisfies traditional internalist epistemic virtues.

Now let's look a little deeper into the predictions made by Lance, Hobart and a proper linear model. While Hobart's unit weight model is usually almost as reliable as the proper model, that doesn't mean they almost always make the same predictions. What it does mean is that when they make different predictions, the proper model is not much more likely to be correct than the unit weight model. We can represent this state of affairs as follows.[7]

Figure 1 about here

We can think of this as a large random sample of the prediction problems Hobart and Lance are likely to tackle with their respective reasoning strategies. When the models make the same predictions ("models agree"), those predictions are (obviously) equally reliable. But when the models make different predictions, the unit weight model is about as reliable as the proper model.

Now let's consider Lance. We know that Lance is less reliable than the proper model. So let's assume that sometimes when the model's prediction is true, Lance's prediction is false (F_1,F_2), and when the model's prediction is false, Lance's prediction is true (T_1,T_2).

Figure 2 about here

The important empirical point established by the SPR findings is that Lance is wrong more often than the proper model ($(F_1 + F_2) > (T_1 + T_2)$).

Now consider Lance's predictions and those of the unit weight model.

Figure 3 about here

Once again, the unit weight model's predictions are more reliable than Lance's predictions ($(f_1 + f_2) > (t_1 + t_2)$). But consider a perfectly possible scenario: The predictions of Lance and the proper model are more alike than the predictions of the unit weight model and the proper model (i.e., $(f_1 + t_1) < (t_2 + f_2)$).

Perhaps an example will help clarify these points. Suppose that after perusing the applications of Smith and Jones, Lance and Hobart disagree about who will be the stronger student. So we're supposing that this is a prediction falling within f_1, t_1, t_2 or f_2. Suppose further that Lance and Hobart decide not to raise the issue of whose prediction is most likely to be <u>true</u>. Instead, they decide to ask whose prediction best satisfies traditional internalist epistemic virtues. Whose prediction has the best reasons in its favor, or best fits the available evidence, or is most coherent with their beliefs? Recall

that according to [3] above, the prediction of a well−constructed proper linear model best satisfies traditional internalist epistemic virtues. So another way to put this issue is: Whose prediction agrees with that of the proper linear model? Notice that it is perfectly possible that the proper linear model would make Lance's prediction (t_2 or f_2). In other words:

4. In the above scenario, Lance's prediction might best satisfy traditional internalist epistemic virtues.

Indeed, whenever Lance and Hobart disagree, Lance will have a very powerful argument for thinking that [4] is true. Lance will have an argument in favor of his prediction and against Hobart's that appeals to (a) evidence that Hobart has intentionally ignored, (b) the relative predictive powers of the cues, which Hobart has also intentionally ignored, or (c) both. Hobart can argue that his prediction is more likely to be <u>true</u>, by appealing to the flat maximum principle and decades worth of completely one−sided evidence for thinking that unit weight models outperform humans. But Hobart has no reply to the argument contending that Lance's prediction better satisfies traditional internalist epistemic virtues. After all, Hobart <u>does</u> ignore evidence, and he <u>does</u> fail to weigh the evidence according to its predictive value. Assuming that Lance and Hobart, like most of the rest of us, do not have a proper model available to decide the question, Hobart cannot defeat Lance's argument.

5. In the above scenario, Lance always has an argument that Hobart can't defeat to the conclusion that Lance's prediction satisfies traditional internalist epistemic virtues better than Hobart's prediction.

The situation described here is paradoxical. The epistemically responsible reasoner will employ a unit weight model (according to [2]), and will reason to Hobart's prediction. But when Lance and Hobart disagree, it is possible that Hobart's prediction is contrary to the belief that best satisfies traditional internalist epistemic virtues (according to [3] and [4]). In fact, the situation here described is so common that it is a virtual certainty that sometimes, the human's prediction (and not the SPR's prediction) is the one that best satisfies traditional internalist epistemic virtues.

6. Sometimes, Lance's prediction really does satisfy traditional epistemic virtues better than Hobart's prediction.

In fact, it is possible that when they disagree, Hobart's prediction is more often than not contrary to the belief that best satisfies traditional internalist virtues; that is, it is possible that $(f_1 + t_1) < (t_2 + f_2)$. What this means is that Lance's predictions might better satisfy traditional epistemic virtues more often (more reliably) than Hobart's predictions. As a result, on any internalist or externalist view of justification, it is possible that Lance <u>knows</u> that his predictions satisfy traditional epistemic virtues better than Hobart's predictions.

Recall our earlier discussion about Peirce's Problem and overconfidence: Intelligent people grant that in the long run, SPRs will be right more often than human

experts, but their (over)confidence in subjective powers of reflection often leads them to deny that in some particular case, the SPR rather than the human will be right. Our discussion here offers a plausible (and perhaps overly generous) reason why people resist SPRs in practice. Using SPRs will sometimes force us to flout our deeply held internalist epistemic convictions. When reasoners reject a SPR, that rejection might well reflect their strong commitment to traditional internalist epistemic virtues. They prefer beliefs that they have overwhelming reason to think has the best reasons in its favor, best fits or coheres with the available evidence, etc. This is not crazy or stupid – far from it. On many views, it's not even irrational. On our view, the rejection of SPRs simply reflects a commitment to faulty epistemic principles. Responsible reasoners will occasionally go out of their way to intentionally flout traditional internalist virtues.[8]

Given the argument in this section, the responsible reasoner will not always reason to the belief she has overwhelming reason to believe best satisfies traditional internalist epistemic virtues; what's more, it is overwhelmingly likely that the responsible reasoner will sometimes adopt a belief knowing full well that it violates such virtues. So let's consider the following principle.

7. If a belief is arrived at by epistemically responsible reasoning, then that belief is epistemically justified.

A number of prominent internalists see a tight connection between justification and deontic notions like responsibility and duty, and so would accept [7] or something like it. For example, Hilary Kornblith offers the following motivation for investigating a responsibility−based concept of justification: "When we ask whether an agent's beliefs are justified we are asking whether he has done all he should to bring it about that he have true beliefs. The notion of justification is thus essentially tied to that of action, and equally to the notion of responsibility" (1983, 34). Laurence BonJour also assumes that the function of epistemic responsibility involves guiding our cognitive endeavors toward the truth. "[O]ne's cognitive endeavors are epistemically justified only if and to the extent that they are aimed at [truth], which means very roughly that one accepts all and only those beliefs which one has good reason to think are true. To accept a belief in the absence of such a reason, however appealing or even mandatory such acceptance might be from some other standpoint, is to neglect the pursuit of truth; such acceptance is, one might say, epistemically irresponsible" (1985, 8).

But [7] presents the internalist with a dilemma. If the internalist accepts this connection between responsibility and justification, then internalism is false. The internalist takes justification to be a matter of a belief having some internalist virtue. But given [5] and [6], sometimes responsible reasoning leads to a belief that one knows does not best satisfy the internalist's epistemic virtue (i.e., one has good reason to think it doesn't, and it doesn't). So according [7], such a belief would be justified; but according to internalism, such a belief would not be justified. So the internalist can't accept [7]. But by rejecting [7], the internalist severs the tight connection between notions of epistemic responsibility and (internalist) epistemic justification. Recall that according to [1], epistemic responsibility is essentially action−guiding: Ignoring non−epistemic consideration, we ought to reason responsibly. By severing the connection with this

prescriptive notion of epistemic responsibility, the internalist is driven to some unpalatable conclusions.

8. Epistemic internalism implies that, ignoring non−epistemic considerations, we sometimes (epistemically) ought to reason to epistemically unjustified beliefs.

9. Epistemic internalism implies that, ignoring non−epistemic considerations, sometimes we (epistemically) ought to reason to beliefs we know are epistemically unjustified.

We believe these implications are intuitively disquieting. But if an internalist accepts them, it is hard to see what motivates an internalist conception of epistemic justification. Why is it important? Epistemology is important and interesting because it addresses the question: How ought we to reason? But the internalist doesn't offer a clear epistemic goal to strive for. For a wide range of reasoning problems, the epistemically responsible internalist is forced to reason to beliefs that by his own favored criteria are unjustified. If that's right, then the question to ask the internalist is: When (epistemically) ought we to reason to justified beliefs? It would appear that the only answer available is: Whenever responsible reasoning happens to hit on them. But if that's so, then why focus so much attention on justification? Responsibility is where the action is.

4. The nature of explanation

Internalism about justification reaches far and wide. For example, theories of explanation tend to depend upon the notion of understanding, and the understanding that an explanation conveys is thought to be justificatory (Trout, unpublished manuscript). The requirement of a sense of understanding may result from an internalist account of justification, an account that states that the determinants of justification are both internal and accessible to the knower. As we have seen, these criteria lead to predictions outperformed by SPRs. Indeed, these "internal and accessible" mechanisms are precisely those responsible for such documented epistemic embarrassments as the overconfidence bias.

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 the psychology of the 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 something like 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 asserts a "fundamental relationship between explanation and understanding." (1983, p.16) Wesley Salmon proposes that scientific understanding is achieved in two ways: by "fitting phenomena into a comprehensive scientific world−picture" (1998, p.77), and by detailing and thereby exposing the "inner mechanisms" of a process (1998, p.77). Michael Friedman claims that the relation of phenomena that "gives understanding of the explained phenomenon" is "the central problem of scientific explanation" (1974, p.189) Philip Kitcher 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" (1981, p.168). James Woodward 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". (1984, p.249) 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.

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, p.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 the important role of 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: "…[I]n 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, p.120) And scientific explanations are no different in this respect; they should "provide a feeling of understanding" (1998, p.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 Ptolemy. Or better yet, 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 evidence 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 narratives, especially dramatic narratives compellingly told, 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 real understanding. Further, the hypothesis about how science works might well fit coherently with all the evidence we know about that we deem relevant. But again, it is a mistake to suppose that responsible reasoning necessarily involves attending 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's 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 (i.e., the representativeness or typicality of these events). 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 clear how to even begin: How do we individuate episodes? What features do we consider in coding them? And since it's impractical to examine all historical episodes, how do we select which ones to consider? These are difficult questions that must at least be addressed, if only in a preliminary way, 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 therefore falls squarely within contemporary naturalistic approaches to the philosophy of science. 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 to draw general conclusions about science, 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. A debiased epistemology of the future

If the superior accuracy of SPRs vindicates a reliablist epistemology, it also points the way to the improvement of methods for acquiring knowledge in the philosophy of science. After all, we would like the methods used in the philosophy of science to be as reliable as the methods used in the sciences it studies, or at least informed by the best science of the time. For example, once we have a more comprehensive cataloging of significant episodes in the history of science, we may be in a position to identify those variables most associated with progressive movements in science. In particular, SPRs can be a source of discipline in the ongoing effort to reduce the known sources of bias, both in and out of science.

Let's now look at the prospective forms that debiasing might take. An inside strategy for debiasing attempts to improve the accuracy of judgment by creating a fertile corrective environment <u>in the mind</u>. 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 which might require that you select a solution that is not intuitively satisfying, but is objectively correct (Kahneman and Lovallo, 1993).

The most prominent of inside strategies is the "consider the opposite strategy". According to one of the groundbreaking studies on debiasing, people "have a blind spot for opposite possibilities" when making social and policy judgments (Lord, Lepper and Preston, 1984). The most effective "inside" strategies urge people to consider alternative hypotheses for the occurrence of the very event that you believe you understand. While it

is perhaps too much to ask that people shoulder technical burdens in lay life here, there is a portable inside strategy that is marginally effective. For any belief that we can hold with undue certainty (e.g., "New York State is the largest state on the Eastern seaboard", "Los Angeles is west of Reno" or, more tragically, "the defendant is guilty beyond a reasonable doubt"), we can follow a simple rule: "Stop to consider why your judgment might be wrong" (Plous 1993, p.228). For example, ask yourself whether, respectively, you have considered South Atlantic states that get less press, the orientation of the U.S., and your confusion over the DNA evidence. When asked to generate pros and cons for a judgment made, Koriat, Lichtenstein and Fischhoff (1980) demonstrated that overconfidence bias was reduced. Indeed, they found that it was the generation of opposing reasons that did all of the bias−reducing work.

The standard assumption, then, is that bias remediation proceeds by exerting willful control over biases, once exposed. This hopeful, "inside" view usually proceeds by assuring us that being aware of our proneness is the first step in correction. Piattelli−Palmarini, for example, tells us that "We will begin to improve ourselves precisely when we can deal with these very abstractions." (1994, p.14) "They are hard to correct spontaneously, but they can, with a little steady work, be put right by anyone who becomes aware of them." (1994, p.15) Optimism about our internal powers always has a ready audience, especially among the Enlightenment hardcore. But once tethered to data, the optimistic view is difficult to sustain in full flora. But now that we know the treachery of subjective judgment, it would be hypocrisy to ignore it or, worse yet, suggest that common sense counteracts this treachery. We now know that general admonitions to concentrate or attend to the evidence does not improve people's performance. Such instructions simply invoke the already defective cognitive routines: "[B]iases in social judgment can be corrected only by a change in strategy, not just by investing greater effort in a strategy that led to biased judgments in the first place." (Lord, Lepper and Preston, pp.1236−1237.)

It is tempting to take heart in the modest success of an inside strategy. However, as hopeful as one might want to be about this finding, it actually provides the first solid evidence inside strategies are local and limited debiasers; their effect is marginal domain−specific. To the extent that "consider−the−opposite" strategies work, they work only for overconfidence and hindsight biases. Moreover, the strategy is difficult to export to a natural setting. This is not a criticism of the ecological validity of the experiments; there is no question that, if you could get people in natural settings to perform the same experimental debiasing task, overconfidence would be reduced. The question is instead whether, as you walk through the day, people will have the discipline, motivation, and concentration required to implement the consider−the−opposite strategy.

It is the ability of outside strategies to rise above the impetuous and interminable seductions of the subjective life that makes them so attractive. This is not to say that inside strategies have no application. In highly structured contexts in which deliberation is mandated and deliberate, as it is on a jury, the social (and other) costs of inside strategies is low. There, we can, and should, consider the opposite. But in other contexts, the inside strategy would make us hopeless, tedious bores, madly excogitating before every substantial remark we make, and after everything everyone else says. Correcting

them whenever the import of their remarks deviate from your calculations, and perhaps equally irritating, confirming the accuracy of your claim whenever you are right. We would lead accurate but lonely lives.

One might suppose that accuracy improves with mere experience in making judgments. People learn various things from experience, no doubt, but they don't appear to learn how to remediate their judgmental distortions. For that, such factors as variability in the environment must be carefully controlled, so that feedback is unambiguous. Tversky and Kahneman (1986) contend that life experience by itself is unlikely to improve judgment performance because:

> (i) outcomes are commonly delayed and not easily attributable to a particular action; (ii) variability in the environment degrades the reliability of feedback...; (iii) there is often no information about what the outcome would have been if another decision had been taken; and (iv) most important decisions are unique and therefore provide little opportunity for learning ...any claim that a particular error will be eliminated by experience must be supported by demonstrating that the conditions for effective learning are satisfied (pp.274−275)

So sheer experience does not seem to produce improved performance; perhaps the acquisition of expertise does. But various studies on expert decision−making shows that simple experience is too complex by itself to allow us to extract subtle theoretical lessons. Experience is no substitute for having either a correct theory or an accurate rule of inference. One might hope that awareness of a problem in judgment leads to correction of a sub−optimal decision strategy. But in each of the above cases, the individual is aware that their judgment is unreliable; they are simply unable to do anything about it on their own. Not surprisingly, then, over the last two decades, research on the nature of bias has demonstrated that bias is not easily counteracted (Fischhoff, et. al., 1977).

In order to correct the structure of scientific theorizing, we must deploy an "outside view", adopting a policy to perform meta−analyses of the literature, for example, even when you think you can extract lessons from eyeballing the history of the field. These policies, like Ulysses's posture toward the sirens, will allow us to accomplish the ends we know are best for us, even when, for all the world, we want to do otherwise. When it comes to health, science policy and the advance of science, a prideful principle of individual judgmental autonomy is no longer benign. Ulysses commended that his crewman "must bind me hard and fast, so that I cannot stir from the spot where you will stand me…and if I beg you to release me, you must tighten and add to my bonds." (The Odyssey)

Does the decisive success of outside strategies imply either that subjective judgment is always unreliable, or that theoretically untutored notions are always scientifically disreputable? No. But it doesn't help. The success of the actuarial approach in the philosophy of science implies a number of lessons. 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 in an intellectual setting that self−consciously endorses the modern scientific culture's attachment to 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?[9]

Notes

1. A common complaint against the SPR findings begins by notes that the whenever humans are found to be less reliable than SPRs, humans are typically forced to use only evidence that can be quantified (since that's the only evidence that SPRs can use). The allegation is that this 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' "human experience", "intuition", "wisdom", "gut feelings" or other distinctly subjective human faculties. Besides the fact that this is an expression of hope rather than a reason to doubt the SPR findings, 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 in that 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.

2. A legitimate worry about SPRs has come to be known as the "broken leg" problem. Consider an actuarial formula that accurately predicts an individual's weekly movie attendance. However, if we knew that the subject was in a cast with 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'd just used the SPR (Goldberg 1968, Sawyer 1966, Leli and Filskov 1984). Our

inclination is to suggest that overriding and 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 SPRs predictions by taking into account such evidence (Swets, Dawes, Monohan 2000, 11).

3. In lay circles, this neglect is understandable. The variables in actuarial formulas reflect years of theoretically arcane research, and their accuracy was established through careful experimental test and statistical analysis. The theoretical knowledge of these findings is not easily digestible by the general public, even a motivated and intelligent public. Moreover, actuarial rules typically require not just knowledge of the values of the variables, but also their functional relations (additive, multiplicative, etc.). So their application often requires patience, discipline and concentration. And a calculator helps.

4. Insofar as this question assumes that there are epistemic considerations that are not reducible to pragmatic or moral considerations, various pragmatists will take this to be a non−sensical question. Given that our aim is to criticize epistemic internalism, we will grant that we are begging questions against pragmatists.

5. Some might deny that epistemic responsibility is essentially action−guiding. Our argument will not suffer much if such objectors replace "epistemic responsibility" with the technical expression "epistemic do−it−iveness" which is (by fiat) essentially prescriptive.

6. In the admissions example, we may assume that the reasoners believe that the cues considered by Lance and by Hobart are predictive. This won't always be the case. Some actuarial formulas include cues and no one has any idea why that cue is predictive of the target property. In such cases, it might not be appropriate to identify the prediction of a proper model with the belief that best satisfies the internalist virtue (depending, of course, on the nature of that virtue).

7. The figures represent simple binary prediction problems (e.g., "Is this patient psychotic or neurotic?"), not more complicated prediction problems (e.g., "Which applicants will be the strongest students?"). Our focus on relatively simple problems does not detract from the general philosophical points we wish to make.

8. A number of internalists have objected to the argument here presented as follows: "It's a mistake to identify the belief that best satisfies internalist virtues with the proper model's prediction. If the improper (unit weight) model is just as accurate as the proper model, that must mean that the extra considerations that the proper model takes into account are irrelevant. And no internalist should be saddled with claiming that a justified belief should be coherent with or should best fit (etc.) <u>irrelevant</u> evidence." This objection is puzzling in some respects, but the simple response to it is that typically (but not always) proper models are more accurate than improper ones. So the extra evidence the proper model considers is not usually completely irrelevant. It's just that as a practical matter, the extra evidence adds so little to the model's accuracy that it is often not worth the trouble.

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

References

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

Ashenfelter, O., Ashmore, D., and Lalonde, R. 1995. "Bordeaux wine vintage quality and the weather, in Chance, 8: 7−14.

Bishop, M. 1999. "Semantic Flexibility in Scientific Practice: A Study of Newton's Optics" in Philosophy and Rhetoric 32: 210−232.

Bishop, M. In progress. "Responsibility Reliabilism".

Bloom, R.F. and Brundage, E.G.: 1947, "Predictions of Success in Elementary School for Enlisted Personnel", in Stuit, D.B. (ed.), Personnel Research and Test Development in the Naval Bureau of Personnel, Princeton University Press, Princeton, pp. 233−261.

BonJour, L.: 1985, The Structure of Empirical Knowledge, Harvard University Press, Cambridge.

Brewer, W., Chinn, C., and Samarapungavan, A. 1998. Explanation in Scientists and Children. Minds and Machines, 8, 119−136.

Carroll, J., et. al. 1982. Law Society Review 17.

Carpenter, R., Gardner, A., McWeeny, P. and Emery, J. 1977. "Multistage scory system for identifying infants at risk of unexpected death" in Arch. Dis. Childh. 53: 606−612.

Dawes, R.: 1971, "A case study of graduate admissions: Application of three principles of human decision making", American Psychologist 26, 180−88.

Dawes, R.: 1979/82, "The robust beautiy of improper linear models in decision making" in Kahneman, D., Slovic, P., Tversky, A. (eds.), Judgment under uncertainty: Heuristics and biases, Cambridge University Press, Cambridge, pp. 391−407.

Dawes, R.: 1994, House of Cards: Psychology and Psychotherapy Built on Myth, The Free Press, A Division of Macmillan, New York.

Dawes, 2000. "A theory of rationality as a 'reasonable' response to an incomplete specification" in Synthese, 1−2: 133−163.

Dawes, R., and Corrigan, B.: 1974, "Linear models in decision making" in Psychological Bulletin 81, 95−106.

Dawes, R., Faust, D. and Meehl, P. 1989. "Clinical versus actuarial judgment" in Science, 243: 1668−1674.

DeVaul, R.A., Jervey, F., Chappell, J.A., Carver, P., Short, B., O'Keefe, S.: 1957, "Medical School Performance of Initially Rejected Students", Journal of the American Meical Association 257, 47−51.

Einhorn, H.J., and Hogarth, R.M.: 1975, "Unit weighting schemas for decision making", Organizational Behavior and Human Performance 13, 172−192.

Faust, D., and Meeh, P. 1992. Using Scientific Methods to Resolve Enduring Questions within the History and Philosophy of Science: Some Illustrations. Behavior Therapy, 23, pp.195−211.

Faust, D. and Ziskin, J. 1988. "The expert witness in psychology and psychiatry" in Science, 241: 1143−1144.

Feldman, R. 1985. "Reliability and Justification" in The Monist, 68: 159−174.

Fischhoff, B., Slovic, P., and Lichtenstein, S. 1977. Knowing with Certainty: The Appropriateness of Extreme Confidence. Journal of Experimental Psychology: Human Perception and Performance, 3, 552−564.

Friedman, M. 1974. Explanation and Scientific Understanding. Journal of Philosophy, 71, 5−19. Reprinted in Theories of Explanation, ed. J. C. Pitt, pp.188−198. New York: Oxford University Press, 1988. Page references are to this reprint.

Gilovich, T.: 1991, How we know what isn't so, The Free Press, A Division of Macmillan, New York.

Goldberg, L.: 1968, "Simple Models of Simple Processes? Some Research on Clinical Judgments", American Psychologist 23, 483−496.

Goldberg, L.: 1970, "Man vs. model of man: A rationale, plus some evidence, for a method of improving on clinical inferences", Psychological Bulletin 73, 422−432.

Golding, J., Limerick, S. and MacFarlane, A. 1985. Sudden Infant Death. Somerset: Open Books.

Goldman, A. 1979. "What is Justified Belief?" in Justification and Knoweldge, George Pappas (ed.). Dordrecht: D. Reidel.

Goldman, A.: 1986. Epistemology and Cognition, Harvard University Press, Cambridge.

Henrion, M., and Fischhoff, B. 1986. "Assessing uncertainty in physical constants" in American Journal of Physics 54: 791−798.

Kahneman, D., and Lovallo, D. 1993. Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. Management Science 39, 1, 17−31.

Kitcher, Ph. 1981. Explanatory unification. Philosophy of Science, 48, pp.507−531. Reprinted in Theories of Explanation, ed. J. C. Pitt, pp.167−187. New York: Oxford University Press, 1988. Page references are to this reprint.

Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980). Reasons for Confidence. Journal of Experimental Psychology: Human Learning and Memory, 6, 107−118.

Kornblith, H.: 1983, "Justified belief and epistemically responsible action" The Philosophical Review 92, 33−48.

Leli, D., and Filskov, S. 1984. Journal of Clinical Psychology 40.

Lord, C., Lepper, M. and Preston, E. 1984. Considering the Opposite: A Corrective Strategy for Social Judgment. Journal of Personality and Social Psychology, 47, 1231−1243.

Lovie, A. D., Lovie, P.: 1986, "The flat maximum effect and linear scoring models for prediction", Journal of Forecasting 5, 159−168.

Lowry, C. 1975. "The identification of infants at high risk of early death" in M.Sc. (Med. Stats.) Report, London School of Hygiene and Tropical Medicine.

Meehl, P.: 1954, Clinical versus statistical prediction: A theoretical analysis and a review of the evidence, University of Minnesota Press, Minneapolis.

Meehl, P. 1986. "Causes and effects of my disturbing little book" in Journal of Personality Assessment 50: 370−375.

Miller, R. W. 1987. Fact and Method. Princeton: Princeton University Press.

Milstein, R.M., Wildkinson, L., Burrow, G.N., Kessen, W.: 1981, "Admission Decisions and Performance During Medical School", Journal of Medical Education 56, 77−82.

Oskamp, S.: 1965, "Overconfidence in Case Study Judgments", Journal of Consulting Psychology 63, 81−97.

Passell, P. 1990. "Wine equation puts some noses out of joint" in The New York Times, March 4, p. 1.

Peirce, C.S. 1878. "Doctrine of Chances" in Writings of Charles Sanders Peirce: A Chronological Edition (Bloomington, Ind.).

Piattelli−Palmarini, M. 1994. Inevitable Illusions. New York: Wiley.

Plous, S. 1993. The Psychology of Judgment and Decision−Making. New York: McGraw−Hill.

Salmon, W. 1998. "The Importance of Scientific Understanding". In Causality and Explanation. New York: Oxford University Press, pp.79−91.

Sawyer, J. 1966. "Measurement and prediction, clinical and statistical" in Psychological Bulletin, 66: 178−200.

Stillwell, W., Barron, F. and Edwards, W. 1983. "Evaluating credit applications: a validation of multiattribute utility weight elicitation techniques" in Organ. Behav. Hum. Perform. 32: 87−108.

Swets, J., Dawes, R., and Monohan, J. 2000. "Psychological science can improve diagnostic decisions" in Psychological Science in the Public Interest 1: 1−26.

Taylor, Shelley. 1989. Positive Illusions: Creative Self−Deception and the Healthy Mind. New York: Basic Books.

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

Trout, J.D. Unpublished manuscript. "The psychology of scientific explanation".

Tversky, A., and Kahneman, D. 1986. Rational Choice and the Framing of Decisions. Journal of Business, 59, S251−S278.

Wedding. 1983. Clinical Neurophychology V 49.

Wittman, M. 1941. "A Scale for Measuring Prognosis in Schizophrenic Patients" in Elgin Papers, 4: 20−33.

Woodward, J. 1984. "A theory of singular causal explanation." Erkenntnis, 21, pp.231−262. Reprinted in Explanation, ed. D. H. Ruben, pp.246−274. New York: Oxford University Press, 1993. Page references are to this reprint.