MICHAEL A. BISHOP

IN PRAISE OF EPISTEMIC IRRESPONSIBILITY: HOW LAZY AND IGNORANT CAN YOU BE?

ABSTRACT. Epistemic responsibility involves at least two central ideas. (V) To be epistemically responsible is to display the virtue(s) epistemic internalists take to be central to justification (e.g., coherence, having good reasons, fitting the evidence). (C) In normal (non-skeptical) circumstances and in the long run, epistemic responsibility is strongly positively correlated with reliability. Sections 1 and 2 review evidence showing that for a wide range of real-world problems, the most reliable, tractable reasoning strategies audaciously flout the internalist's epistemic virtues. In Section 3, I argue that these results force us to give up either (V), our current conception of what it is to be epistemically responsible, or (C) the responsibility-reliability connection. I will argue that we should relinquish (V). This is likely to reshape our epistemic practices. It will force us to alter our epistemic judgments about certain instances of reasoning, to endorse some counterintuitive epistemic prescriptions, and to rethink what it is for cognitive agents to be epistemically responsible.

Imagine you're serving on an undergraduate admissions committee. There are various strategies you might adopt for predicting the future academic success of applicants. You might review all the available relevant evidence, GPA, test scores, letters of recommendation, quality of high school, etc., try to make the judgments that best fit your evidence and then spend hours wrangling with colleagues. Less onerous strategies are possible. For example, you might predict randomly, or on the basis of a single piece of evidence, like applicants' GPA's or test scores or the weight of their dossiers. Among the feasible reasoning strategies, which is the most epistemically responsible? Given what most academics do when faced with gatekeeping tasks, the answer seems clear enough – the conventional, labor-intensive strategy. But this answer is false. My aim in this paper is to explain why. Along the way, I'll argue that once we understand why the conventional answer is false, we will have to rethink the epistemic judgments and prescriptions we endorse, as well as the nature of our central epistemological concepts.

Synthese **122:** 179–208, 2000.

The notion of responsibility that has dominated analytic epistemology involves two distinct ideas. The first idea sets out what it is to be epistemically responsible.

(V) To be epistemically responsible is to display in one's reasoning the virtue (or virtues) epistemic internalists take to be central to warrant or justification, e.g., coherence, having good reasons, fitting the evidence.

Epistemic internalists believe "that what determines whether a belief is warranted for a person are factors or states in some sense internal to that person" (Plantinga 1993, 5). Different epistemologists will defend different virtues as being essential to warrant or justification. I intend to remain defiantly vague on this point. I have no dog in that fight. But the tension I hope to identify in our understanding of epistemic responsibility does not depend on the details of how one might explicate the internalist's virtues. It will arise for any plausible way of making (V) more precise that does not reduce the relevant virtue to a reliabilist one (e.g., to using reliable belief-forming mechanisms or reliable reasoning strategies). A responsible reasoning strategy is one that would be adopted by a cognitive agent who displays epistemic virtues like coherence, fitting the evidence, having good reasons, etc. So the reason we believe that the most responsible way to approach the admissions problem is to use the conventional, labor-intensive strategy is that this is the best way for us to end up with the prediction that (take your choice) best fits the available evidence, makes ones belief-system most coherent, has the best reasons in its favor, etc.

The second idea implicit in our notion of epistemic responsibility becomes apparent when we ask why epistemic responsibility is important. Or, more to the point, why is it important for cognitive agents to be epistemically responsible? Its value is largely (and perhaps primarily) instrumental. Epistemic responsibility is valuable because of its special relation to truth. For the admissions problem, if we adopt the most responsible strategy, we maximize our chances of making reliable predictions. Reliable judgments typically afford us the best opportunity to effectively intervene in the world to achieve our aims; in this case, if we use the labor-intensive strategy, we'll tend to admit the best students. What I will call *the consilience assumption* says that epistemic responsibility has a special connection to truth.

Before trying to spell out this connection, we should note that systematic consilience failure is very old news to philosophers. The philosophical literature is full of examples of reliability and epistemic responsibility going their separate ways. The most notable example is skepticism. If we live in a skeptical world where our experiences are caused by a powerful demon, intent on deceiving us, a perfectly responsible cognitive agent will have systematically false beliefs. A second well-known example of consilience failure involves the traditional objection to coherence theories of justification. According to this objection, a set of maximally coherent beliefs might be so breathtakingly unreliable that it would be absurd to count them as justified. Here we have responsibility (as coherence maximization) without reliability. Yet a third kind of consilience failure – this time, reliability without responsibility – is also found in the literature. These are typically put forward as objections to reliabilist theories of justification. One version of the objection posits a perfectly reliable clairvoyant. We are supposed to conclude that even though the clairvoyant's predictions are perfectly reliable, it would be highly epistemically irresponsible, and thus unjustified, to believe them.

Even though the philosophical literature is full of examples of *possible* consilience failures, the assumption that there is *as a matter of fact* an intimate connection between reliability and responsibility survives. This is so even for philosophers who are internalists about responsibility and justification. For example, Feldman and Conee argue that doing one's epistemic duty "is the epistemically best way to use one's evidence in trying to believe all and only the truths one considers" (Feldman and Conee 1985, 20). 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 (Kornblith 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*" (BonJour 1985, 8). The consilience assumption survives because except for the village skeptic, no one *really* believes that we live in a world in which consilience systematically fails. We believe that systematic consilience failures would require powerful evil demons, perfectly reliable clairvoyants, epistemically serendipitous brain lesions, and other creatures quirky and contrived. In our everyday lives, when pondering vacation plans, investment strategies or what to have for dinner, most of us manage to reckon reliably and responsibly while putting aside worries about dreamers, demons and madmen. Of course, that's not to say that in the real world consilience is guaranteed. Sometimes guesses are lucky and clairvoyants insightful. But the consilience assumption that most philosophers would embrace can take that fact into account.

(C) In normal (non-skeptical) circumstances and in the long run, being *more* epistemically responsible tends to lead to *more* reliable judgments. The more responsible a reasoning strategy, the more reliable it will tend to be, and the most responsible reasoning strategy will typically be the most reliable.

Insofar as epistemic responsibility is instrumental, the consilience assumption – the idea that being epistemically responsible is more truth-conducive than being epistemically irresponsible – is built right into it.

The importance of both (C) and (V) is evident in our pragmatic decision-making about what reasoning strategies to adopt in particular situations. For example, the pragmatic calculation typically begins by noting that some strategy (usually a labor-intensive one) is the epistemically responsible one (V), and it pays off by optimizing our chances of making reliable judgments (C). But the price of epistemic responsibility is paid for in the coin of time and computational resources. So on pragmatic grounds, no real cognitive agent can be (or should try to be) *maximally* epistemically responsible all the time. Some problems are intractable. Others are just not worth the trouble. So we might often adopt a less-than-perfectly-responsible strategy, and insofar as we are irresponsible, our judgments are likely to be less reliable than they might have been (C). In the case of the admissions problem, being epistemically irresponsible will cost us in terms of accepting relatively poorer students. That cost might be worth paying. But it's still a cost.

Most of this paper is devoted to setting up and arguing for the following thesis: For a wide range of real-world problems, there is no tractable reasoning strategy that is "responsible" according to both (V) and (C). I will suggest that this result has interesting and counterintuitive implications for the epistemic judgments we make, the epistemic prescriptions we endorse and the epistemological concepts we employ. The paper is organized as follows. Sections 1 and 2 review evidence showing that certain reasoning strategies are more reliable than (or as reliable as) others for certain types of problem. What's surprising is that the most reliable reasoning strategies are not the ones we would deem, according to (V), "responsible". In Section 3, I argue that these reliable reasoning strategies do not display the virtue (or virtues) epistemic internalists take to be central to warrant or justification, and I consider the implications of this result. The most direct implication is that for a wide range of real-world reasoning problems, we must give up either (V) or (C). We must give up our current ideas about what it is for a cognitive agent to be epistemically responsible, or we must give up consilience, the idea that greater epistemic responsibility tends to lead to greater reliability (or at least, that it doesn't degrade reliability). Given this choice, I argue that we should retain consilience and revise our conception of what is involved in being epistemically responsible. This will force us to dramatically revise some of our most firmly held judgments about what counts as epistemically responsible reasoning. It will also force us to revise our idea of what virtues are involved in epistemic responsibility. Finally, I suggest that these results render very attractive a reliabilist conception of epistemic responsibility, one that is exhausted by some appropriately articulated version of the consilience assumption.

1. APPARENT CONSILIENCE FAILURES: HOW LAZY CAN YOU BE?

In this section and the next, I will report on apparent consilience failures that require no improbably exceptional reasoners and no systematically bizarre environments. They are extraordinary precisely because there is nothing *else* extraordinary about them. They involve no one that might attract fans to a Coney Island freak show – no supernatural gifts, amazing maladies or coherent delusions. Nor will the landscape be haunted by evil demons, kidnapping brain scientists or fake barns. The examples feature reasoning strategies that adults of average intelligence can and sometimes do use. And they will center on prediction problems people face every day. Let's begin by returning to the admissions problem. As a member of the admissions committee, you dread the idea of pouring over all the dossiers, and you're appalled at the prospect of haggling with argumentative colleagues about which dossiers are best. Is there an easier and more reliable way to make these predictions? Absolutely.

1.1. *Expert vs. Actuarial Predictions*

Paul Meehl's classic book, *Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Literature*, suggests an alternative way to solve the admissions problem. Meehl compared the predictions made by human experts (clinical prediction) against actuarial (statistical) predictions. He defines an actuarial prediction as a purely mechanical procedure, in which

> … the prediction is arrived at by some straightforward application of an equation or table to the data. … The defining property is that no judging or inferring or weighing is done by a skilled clinician. Once the data have been gathered from whatever source and of whatever type, the prediction itself could be turned over to a clerical worker. (Meehl 1954, 15–6)

To get some sense of what actuarial equations are like, consider the instructions Meehl might give to the clerical worker who has been assigned to make actuarial predictions for the admissions problem: For each applicant, (1) take each predictor cue and turn it into a number in accordance with a particular function; (2) multiply each cue number by a weight; (3) add the weighted cue numbers; (4) the sum gives the prediction for that applicant in accordance with some function. In order to get the weights, we begin with a data set that contains information about how the cues and the target property are related. For the admissions problem, this would consist of data that showed the relationship between students' GPA, test scores, etc. (the cues) and their academic performance in college (the target property). Weights are chosen that best fit the available data: they optimize the relationship between the weighted sum of the cues (the actuarial prediction) and the target property in that data set. Call this kind of actuarial equation a *proper linear model* of the admissions problem (Dawes 1979/82, 391).

Which is more reliable when based on exactly the same evidence, actuarial or clinical prediction? Meehl reported on 20 experiments. In every non-ambiguous case, the actuarial predictions were more reliable. Since the publication of Meehl's book, "[t]he bulk (in fact, all) of the literature … supports his generalization about proper models versus intuitive clinical judgment" (Dawes 1979/82, 394). When based on the same evidence, actuarial models are more reliable than human experts for problems of social prediction – not in most studies, but in every study. For example, "[i]n predicting the success of electroshock therapy, a weighting of marital status, length of psychotic distress, and a rating of the patient's 'insight' into his or her condition outperformed one hospital's medical and psychological staff members. In predicting criminal recidivism in several settings, past criminal and prison record outperformed expert criminologists" (Dawes 1994, 83).

One might suspect that human experts do worse than actuarial formulas because they are restricted to the sort of objective information that can be plugged into a formula – they are unable to use their subjective "intuition". Whenever this possibility has been tested, however, it has been found to be false. For example, when human experts and actuarial formulas are given the same evidence, and then humans get more information in the form of unstructured interviews, clinical prediction is *still* less reliable than actuarial prediction. In fact, the unstructured interviews actually make matters worse. They degrade the reliability of human prediction (Bloom and Brundage 1947; DeVaul et al. 1957; Oskamp 1965; Milstein et al. 1981). Another common response to these findings is to suppose that a combination of clinical and actuarial judgment would yield better results than either type of judgment alone. In fact, that's not true either. When trained subjects were given the results of the actuarial formula to help them make their judgment, and told that such formulas typically outperform human predictors, they still did worse than the formula (Goldberg 1968; see note 2).

These are fascinating and too-little-known findings. They have moral and pragmatic implications about how certain social practices and institutions ought to operate – implications that many people find disturbing (Dawes 1994). But for our purposes, the most important feature of the proper strategy (i.e., using a proper linear model to solve the admissions problem) is that it would be extremely laborious. It would involve (1) collecting data about the correlations between the cues and the target property, (2) hiring a good statistician to derive the proper actuarial formula, and (3) putting the formula into a form that can be easily computed. Without access to lots of data, a good statistician and a computer, the proper strategy would probably involve more work than the conventional labor-intensive strategy. For most of us faced with gatekeeping problems, these results won't help ease our burdens.

1.2. *Experts vs. Virtual Experts*

A *proper* linear model assigns weights to predictor cues (GPA, test scores) so as to optimize the relationship between those cues and the target property (academic success) in a data set. Once we understand that these models are built to best fit the available data, perhaps we shouldn't be surprised that they outperform human experts. What about *improper* linear models, models that are not constructed so as to best fit the available data? How do they stack up to expert human prediction? Goldberg (1970) asked 29 clinical psychologists to predict, only on the basis of a Minnesota Multiphasic Personality Inventory (MMPI) profile, whether a patient would be diagnosed as neurotic or psychotic. He then constructed 29 proper linear models that would mimic each psychologist's predictions. The predictor cues consisted of the MMPI profile; the target property was the psychologist's predictions. Weights were assigned to the cues so as to best fit *the psychologist's predictions* about whether the patient is neurotic or psychotic. A *bootstrapping* model is a proper linear model of a human's predictions, but it is an improper linear model of the target property – in this case, the patient's condition.

One might expect that the bootstrapping model will predict passably well. It is built to mimic a fairly reliable expert, so we might expect it to do nearly as well as the expert. In fact, *the mimic is more reliable than the expert*. Goldberg found that in 26 of the 29 cases, the bootstrapping model was more reliable in its diagnoses than the psychologist on which it was based! (For other studies with similar results, see Wiggins and Kohen 1971; Dawes 1971.) This is surprising. The bootstrapping model is built to ape an expert's predictions. And it will occasionally be wrong about the expert. But when it is, it's more likely to be right about the target property!

The bootstrapping result suggests a strategy for solving the admissions problem and problems of social prediction generally (Goldberg 1970). Construct a proper linear model of your own predictions, and then substitute the model's judgments for your own. This will be more reliable than the traditional labor-intensive strategy, and less onerous than the proper strategy. The proper strategy requires a reliable data set that shows the relationship between cues (GPA, test scores) and the target property (future academic performance). To construct the bootstrapping model, you need no information about the target property. You only need a data set showing the relationship between the cues and your own predictions. To construct the model, you'll still need a statistician and a computer. Once it's done, though, you can use it to make your admissions decisions for years to come. While the bootstrapping strategy is less work than the proper strategy, it's not clear that it's less work than the conventional labor-intensive strategy. The laziest among us might hope for better.

1.3. *The Flat Maximum Principle*

Why are bootstrapping models more reliable than the humans on which they are based? The early hypothesis was that bootstrapping models somehow capture the underlying reliable policy humans use; but since they are not subject to extraneous variables that degrade human performance, the models are more accurate (Bowman 1963; Goldberg 1970; Dawes 1971). This is a relatively flattering hypothesis, in that it grants us an underlying competence in making social predictions. But Dawes and Corrigan (1974) showed that this hypothesis is false. They took five bootstrapping experiments and for each one constructed a *random* linear model, one that assigned random weights to predictor cues (except for sign, i.e., if a cue was positively [negatively] correlated with the target property, the random weight assigned to it was always positive [negative]). The random models were about as reliable as the bootstrapping models, and were therefore more reliable than humans. In cases of social prediction, we would do better to rely on a random linear model than on an epistemically responsible human expert!

So why do improper models outperform human experts? The answer is: Because the proper models do. There is an analytic finding in statistics called the flat maximum principle (Lovie and Lovie 1986). Statistical technicalities aside, here's what it says.

(FM) If a prediction problem is such that no proper model will be especially reliable and if the cues are reasonably predictive and somewhat redundant, then the reliability of a linear model's predictions will not be particularly sensitive to what weights are assigned to the cues (except for their sign).

These conditions will typically be satisfied for problems of social prediction. For example, in the case of the admissions problem, even the best models are not exceptionally reliable, the best cues (GPA, test scores) are reasonably predictive, and they are somewhat redundant (e.g., people with higher GPA's tend to have higher test scores). So for problems of social prediction, improper models will be about as reliable as proper models. And since humans do considerably worse than proper models, it follows that improper models will outperform humans too.

Among the improper linear models, there is one that tends to be a bit more reliable than the others. Unit weight models assign equal weights to (standardized) predictor cues, so that each cue has an equal "say" in the final prediction. Given the flat maximum principle, it's no surprise that the unit weight model is about as reliable as the proper model. In fact, Einhorn and Hogarth (1975) have shown that there are not uncommon situations in which the improper unit weight models are *more* reliable than the proper models. As Paul Meehl has said, "In 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 succinctly state the cash value of these results: To be more reliable than epistemically responsible humans in the social arena, "the whole trick is to know what variables to look at and then know how to add" (1974 105).

Given the results we've reviewed, it is no surprise that "[i]n predicting academic performance ... a simple linear weighting of high school rank and aptitude test scores outperformed the judgments of admissions officers in several colleges" (Dawes 1994, 83). In absence of the results we've reviewed, this is a huge surprise; so much so that a typical reaction to them is disbelief, especially among the human experts who are consistently outperformed by the improper models! So when it comes to the admissions problem, in the long run, you will *outperform* (in terms of reliability) the conventional labor intensive strategy by simply adding each students' high school rank (out of 100) to their test score rank (out of 100) and then predicting that the candidates with the highest totals will be the best students. Here at last, we've found a solution to the admissions problem that will satisfy the laziest of cognitive agents. You can turn it over to a clerical worker – or better yet, a graduate student. And if that weren't virtue enough, in the long run, no tractable strategy will be more reliable.

The take-home lesson of the above results is *not* that we can get a purely mechanical procedure to outperform humans. After all, the proper and improper models predict reliably because skilled humans have set them up to predict reliably. The take-home lesson of these results is that for problems of social prediction, *we can increase our reliability by rejecting labor intensive strategies that intuitively strike us as epistemically responsible and using instead simple but intuitively improper methods of integrating the information we have*. The unit weight model is an irresponsible but very reliable strategy (or IRS) for making social predictions. Four features of this IRS are especially noteworthy. First, it has reasonably transparent conditions of appropriate application; that is, the conditions under which using it would be appropriate are fairly easy to spot. Second, the IRS is tractable; it is relatively easy for adults of average intelligence to use. Unlike some highly responsible reasoning strategies, e.g., Bayesian predictive strategies, this IRS can be used by people who don't possess a computer or lots of training in statistics. Third, as long as the IRS is appropriately applied, it is highly reliable. It is about as reliable as intractable ideal strategies, and it is more reliable than humans attempting to be responsible. And fourth, the use of the IRS violates our considered views about what it is to be epistemically responsible (an argument for this is in Section 3.1). Of course, human reasoners don't use this particular IRS when faced with social prediction problems. If we did, we wouldn't be consistently outperformed by proper and improper linear models. Section 2 will introduce more IRS's, and more examples of apparent consilience failures. What is different about the IRS's to be considered in Section 2 is that, unlike unit weight models, there is evidence that human reasoners are sometimes naturally inclined to use them.

2. CONSILIENCE FAILURES: FAST AND FRUGAL HEURISTICS

Proper linear models outperform experienced, responsible humans. Pessimistic results like these about human reasoning should sound familiar to those who have followed, even casually, findings in the psychology of reasoning over the past three decades. The heuristics-and-biases literature is full of examples in which humans make systematic mistakes in deductive and probabilistic reasoning (for excellent reviews, see Tversky and Kahneman 1974; Nisbett and Ross 1980; Gilovich 1991). The explanation for the mistakes is that humans typically employ relatively simple heuristics rather than rules of deductive logic or probability when reasoning. And while these heuristics are often reliable, they can lead to systematic errors.

> In making predictions and judgments under uncertainty, people do not appear to follow the calculus of chance or the statistical theory of prediction. Instead, they rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors. (Kalineman and Tversky 1973/82, 48)

An ideally responsible cognitive agent would "follow the calculus of chance or the statistical theory of prediction" and wouldn't violate laws of logic. We, on the other hand, rely on heuristics that are fairly responsible and fairly reliable. We're like the C student who cuts corners and is doing well to scrape by. Many have chafed against the apparent pessimism of the heuristics-and-biases program. L. J. Cohen, for example, denies that empirical results can show that our reasoning competence is less than ideal (Cohen 1981).

The IRS results so far canvassed seem to add to the pessimism: Responsible expert humans aren't even as reliable as random linear models. But notice that the pessimists and the optimists rely on the assumption that our current conception of epistemic responsibility makes consilience true. The pessimists say we're irresponsible and this leads to lots of error. The optimists respond that we're not irresponsible and we don't make (so many) errors. But the IRS results actually suggest a third, and very optimistic possibility. Perhaps we are irresponsible (given [V], our current conception of what it is to be responsible) and nonetheless we're about as reliable as a "responsible" cognizer! We know humans use simple, less-than-ideally responsible reasoning strategies (heuristics); and we know that for certain kinds of problem, simple, less-than-ideally responsible reasoning strategies (heuristics) can be as reliable as, or even more reliable than, proper ones. For at least some of the problems we encounter in the real world, maybe the simple, improper heuristics we use are about as reliable as more expensive "ideal" reasoning strategies. So instead of being like the dim C student, maybe we're like the crafty student who gets A's without working terribly hard. This is part of the proposal made by Gerd Gigerenzer and the ABC group (Gigerenzer and Goldstein 1996; Goldstein and Gigerenzer, 1999; Gigerenzer et al. 1999; Gigerenzer et al. 1999a). My aim in this section will be to review some of their fascinating and suggestive results concerning fast and frugal heuristics.

The heuristics-and-biases program and the flat maximum results are well established. My discussion might lead one to think that the fast and frugal heuristics program is merely an extension of work that has gone before. Such an impression would be mistaken. The work of Gigerenzer and his colleagues springs from different theoretical commitments, normative assumptions and methodologies. Exploring the nature and depth of these differences would take us away from our concerns. But one difference is especially noteworthy for our purposes. The heuristics-and-biases program takes the rules of logic and probability to be a (though not the only) standard against which human performance is to be evaluated. A reasoner who tries to solve a problem and doesn't get the answer sanctioned by logic or probability has made an error. On the other hand, the *primary* test (and ultimately perhaps the only test) Gigerenzer and company set for a reasoning strategy is how reliable it is in real-world environments. This aim opens up the possibility of discovering reliable reasoning strategies that violate canons of epistemic responsibility – in other words, IRS's.

2.1. *Take the Best*

In Section 1, we considered prediction problems in which we assumed we were blessed with very good information, e.g., we knew *exactly* what an applicant's high school rank and test scores were. In many real-life situations, however, we don't have access to such fine-grained information. Gigerenzer et al. (1999a) take on prediction problems in which only coarse-grained data is available. The problem is to predict which of two cities has the higher homelessness rate. The cues are rent control, vacancy rate, temperature, unemployment, poverty and public housing. Vacancy rate and public housing are negatively correlated with homelessness, while the other cues are positively correlated with it. The data are accurate but binary. One cue is naturally binary (rent control) and the others are coded in terms of whether the city is above or below the median on the particular cue. A "1" is assigned to a city when the cue suggests a higher rate of homelessness, otherwise a "0" is assigned. When one object has a value of "1" on a cue and the other does not, then the cue discriminates between the objects. So, for example, the cue *rent control* discriminates between Los Angeles and Chicago (see chart 1). Suppose we use a cue to make a prediction about homelessness rates only if it discriminates between cities; otherwise, no decision is made. We can define the validity of that cue as the number of correct inferences it leads to divided by the total number of inferences (correct and incorrect) it leads to. In Chart 1, the cues are ordered in terms of their validities. So for example, rent control has the highest cue validity. When rent control discriminates between two of the

City (homeless per million)		Rent control	Vacancy Rate	Temp.	Unemp.	Poverty	Public housing
Los Angeles (10,526)		1	1	1	1	1	1
Chicago (6,618)		0	1	0	1	1	1
New York (5,024)		1	1	1	1	1	0
New Orleans (2,671)		0	0	1	1	1	0

Chart 1. Adapted from Gigerenzer et al. (1999a).

cities, the city assigned "1" has the higher homelessness rate 90% of the time.

Now we are in a position to define a fast and frugal heuristic, *Take the Best* (Gigerenzer et al. 1999a). Take the Best (TB) begins with the cue with the highest validity. If the cue discriminates, then TB predicts that the object assigned "1" has the higher value; if it doesn't discriminate, then TB continues on to the cue with the next highest validity. And so on. If no cue discriminates, TB chooses one of the objects randomly. Unlike the various proper and improper models we have examined to this point, TB does not integrate different lines of evidence. It is a one-reason decision maker. It predicts that Los Angeles has a higher homeless rate than Chicago solely on the basis of the rent control cue. It ignores the other five cues. And in this case, TB is right. (Of course, TB will sometimes have to search through more than one cue, but it does not try to integrate them in making a judgment.) Gigerenzer and Goldstein (1996) call Take the Best a fast and frugal heuristic. Fast, because it processes information simply (without integrating different lines of evidence), and frugal because it uses little information (especially compared to models that integrate all the different pieces of information).

Just how reliable is Take the Best? Gigerenzer et al. (1999a) tested it against a multiple regression model and a unit-weight model. There are 1225 city pairings of the 50 largest U.S. cities, and each model made 1225 predictions about which of the pair had the higher homelessness rate. The cue validities for TB, the weights for the multiple regression formula, and the cue's direction (positively or negatively correlated with target property) for the unit-weight formula all came from the entire data set (so test set = training set). When it comes to fitting data, TB is about as reliable as the others. (The hit rates may seem low, but that's because for any algorithm

Strategy		Average number of cues looked up	% correct for data fitting	% correct for prediction
Take the best		2.4	69	63
Unit weight		6	66	58
Multiple regression		6	70	61

Chart 2. Adapted from Gigerenzer et al. (1999a).

Strategy		Average number of cues looked up	% correct for data fitting	% correct for prediction
Take the best		2.4	76	71
Unit weight		7.4	73	70
Multiple regression		7.4	78	67

Chart 3. Adapted from Gigerenzer et al. (1999a).

that must guess whenever the cities have the same cue profiles, the expected upper limit on accuracy is 82%.) TB was much less expensive, looking up an average of 2.4 cues per prediction, whereas the other two needed to consider all six cues in order to make a prediction.

To test the predictive ability of the models, Gigerenzer, et al. (1999a) randomly split the cities into 2 groups of 25, trained up the models on one half of the cities, and then used the models to make predictions on all pairings of the other 25. They randomly split the cities and tested each of the models 1,000 times. Surprisingly, TB outperforms even the multiple regression equation (Chart 2). One's first reaction, and maybe second and third reactions, is that there must be something fishy about this example. So Gigerenzer, et al. tested the models on 20 real world data sets.

In order to make our conclusions as robust as possible, we also tried to choose as wide a range of empirical environments as possible. So they range from having 17 objects to 395 objects, and from 3 cues to 19 cues. Their content ranges from social topics like high school drop-out rates, to chemical ones such as the amount of oxidant produced from a reaction, to biological ones like the number of eggs found in individual fish (manuscript, 9).

The results once again show that when the test set is identical to the training set, TB is almost as accurate as multiple regression, and that when the test set is not identical to the training set, TB is more accurate than the others (Chart 3).

When is Take the Best likely to outperform more expensive and sophisticated algorithms? According to Gigerenzer et al. (1999a), it has been proved that when information is *scarce*, TB will be on average more reliable than the unit-weight model (Martignon et al. manuscript). (Information is scarce when there aren't enough cues to provide each object a

Strategy		& correct for city population	% correct for homelessness	% correct for fish fertility	% correct for prof's salaries
Take the best		74	69	73	80
Bayesian		76	77	74	84

Chart 4. Adapted from Gigerenzer et al. (1999a).

Strategy		& correct for city population	% correct for homelessness	% correct for fish fertility	% correct for prof's salaries
Take the best		72	63	73	80
Bayesian		74	65	75	81

Chart 5. Adapted from Gigerenzer et al. (1999a).

unique profile. In other words, if there are *n* cues and more than 2^n objects, then information is scarce [when information is binary].) TB will do as well as a weighted linear model when the information is redundant. For example, take the cue with the highest validity. Suppose that in a proper linear model, that cue is assigned a weight that is greater than the sum of all the weights assigned to all the other cues; and so too for the cue with the next highest validity, and so on. In this case, the predictions of TB and the proper model will be identical. Gigerenzer et al. (1999a) argue that since "[s]carceness and redundancy of information are characteristics of information gathered by humans" (ms, 14), it shouldn't be surprising that in real-world environments, TB performs about as well as the very sophisticated and computationally expensive models.

Some might object that Take the Best has not been tested against a truly ideal predictive model. How would TB do against a Bayesian predictor? For problems of social prediction, Bayesian methods are extremely complex and computationally very expensive. Such methods are certainly beyond the unaided capabilities of even quite extraordinary human reasoners. Nonetheless, Gigerenzer et al. (1999a) tested TB against a Bayesian network in four different real-world environments: "Which of two German cities has the higher population? Which of two U.S. cities has a higher homelessness rate? Which of two individual Arctic female charr fish produces more eggs? Which of two professors at a Midwestern college has a higher salary?" (manuscript, 18). For data-fitting problems, the Bayesian network outperformed TB by between 1% and 8%; and the Bayesian network was consistently close to perfection (Chart 4). But for prediction problems (training and test sets were not identical), the differences closed dramatically. TB was within 1% to 2% of the Bayesian network (Chart 5).

These are jaw-dropping results. The full explanation for the success of Take the Best is not yet in. There are many possibilities to investigate. For example, one potential explanation for the relative success of TB is that the data were represented in a way that was most friendly to it: the binary representation of data was a disadvantage to the more complex strategies. On the other hand, how would a modified version of Take the Best (one that could make finer discriminations among the data) do against a Bayesian network or a multiple regression equation? We don't yet know.

Do humans rely on something like Take the Best? We don't know for sure, but there is a bit of positive evidence. Payne et al. (1988, 1993) suggest that people favor strategies like TB when they are under time constraints. Regardless of the ultimate psychological plausibility of TB, however, for our purposes, the important point is that TB is a reasoning strategy that appears to be highly reliable for a very wide range of prediction problems where the data are coarsely tuned. For such problems, TB is almost as reliable as Bayesian predictive models, which for all practical purposes are intractable for human reasoners. Further, it is probably more reliable than other highly sophisticated and computationally expensive predictive models.

2.2. *The Recognition Heuristic: How Ignorant Can You Be?*

So far, we have assumed that the responsible and irresponsible reasoning strategies work with the same evidence. But the truly irresponsible epistemic agent would want to get away with less. In fact, there is at least one well-defined situation in which the relatively ignorant epistemic agent can do as well as – or better than – his better informed and responsible colleague. Consider: Which city has more inhabitants, San Diego or San Antonio? U.S. students answered this question correctly 62% of the time. German students, on the other hand, answered the question correctly 100% of the time (Goldstein and Gigerenzer, 1999). Even if we grant the superiority of German education, it's hard to believe German students know more about U.S. geography than U.S. students. Goldstein and Gigerenzer (1999) investigated this result. They took the 22 largest cities in the U.S., randomly paired them, and asked U.S. students to pick the larger (in terms of inhabitants). Then they took the 22 largest German cities, randomly paired them, and asked the students again to pick the larger. The U.S. students did better on the German cities (median 71% versus median 73%). And when Goldstein and Gigerenzer ran this same experiment on German students, they found that the Germans were more accurate on the U.S. cities. They call this the *less-is-more effect*: Under certain circumstances less knowledge can yield more reliability.

What explains the less-is-more effect? Goldstein and Gigerenzer hypothesize that when subjects are somewhat ignorant about a subject, it allows them to employ the *recognition heuristic*: If S recognizes one of two objects but not the other, and S believes that recognition correlates positively (negatively) with the criterion, then S can infer that the recognized object has the higher (lower) value. So consider again the San Diego vs. San Antonio problem. The German students tended to recognize the former city but not the latter, so they used the recognition heuristic and inferred (correctly) that San Diego was larger. The U.S. students recognized both cities and so did not use the recognition heuristic; they made a judgment on the basis of the knowledge they had about the respective cities. In the case of San Diego vs. San Antonio, the recognition heuristic was more reliable. Maybe too much learnin' *can* be a bad thing.

Call the probability of getting a correct answer when one object is recognized and the other is not the *recognition validity*. Call the probability of getting a correct answer when both objects are recognized the *knowledge validity*. The less-is-more effect occurs when the recognition validity is greater than the knowledge validity. Goldstein and Gigerenzer offer an imaginary example that makes clear how this might work.

> Imagine . . . [three brothers] have to take a quiz about German cities at school. The quiz consists of randomly-drawn, two-alternative questions about population sizes of the 50 largest German cities. The youngest brother is ignorant, and has never even heard of Germany (not to speak of German cities) before. The middle brother is savvy, and recognizes 25 of the 50 largest cities from what he has overheard from day to day. The cities this middle brother recognizes are larger than the cities he does not recognize in 80% of all comparisons, that is, the recognition validity is 0.8. The eldest brother is quite the scholar and has heard of all of the 50 largest cities in Germany. When any of the brothers recognizes both cities in a pair, he has a 60% chance of making the correct choice, that is, [the knowledge validity] is 0.6. (Goldstein and Gigerenzer 1999, 45)

On the assumption that the middle brother uses the recognition heuristic whenever he can, the younger brother will perform at a chance level (50% accurate), while the oldest brother will get 60% correct. But the middle brother makes the most accurate inferences. When he recognizes neither, he guesses and gets 50% right; when he recognizes both, he gets 60% right; and when he recognizes just one, he gets 80% right.

Whenever the older and middle brothers disagree about which city is larger, we think that if both were epistemically responsible, the middle brother would defer. A responsible reasoner will defer to someone who is more knowledgeable about a subject. But in this case, at least, that is not the most reliable strategy. Further, if the middle brother decided to become as well-informed about Germany as his older brother before taking this quiz, we would think this an epistemically responsible course of action. But by becoming well-informed, he would degrade his predictions. In this case, epistemic indolence pays.

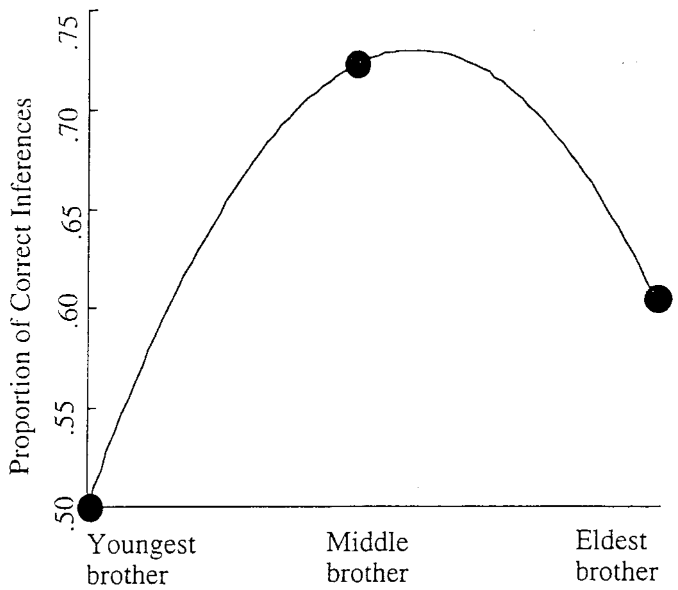

Chart 6. Adapted from Goldstein and Gigerenzer (1999, 46).

3. RETHINKING RESPONSIBILITY

The IRS results put pressure on a wide range of internalist notions of epistemic responsibility. In the introduction, I noted that this notion of epistemic responsibility involves two distinct ideas. The first sets out what it is for an agent to be epistemically responsible.

(V) To be epistemically responsible is to display in one's reasoning the virtue (or virtues) epistemic internalists take to be central to warrant or justification, e.g., coherence, having good reasons, fitting the evidence.

The second idea implicit in our notion of epistemic responsibility is consilience.

(C) In normal (non-skeptical) circumstances and in the long run, being *more* epistemically responsible tends to lead to *more* reliable judgments. The more responsible a reasoning strategy, the more reliable it will tend to be, and the most responsible reasoning strategy will typically be the most reliable.

In Section 3.1, I argue that the IRS results force us to choose between (V) and (C). We can retain our view's about what it is for an agent to be epistemically responsible (V), in which case the IRS results force us to give up consilience (C). Or we can retain consilience (C), but revise our views about what is involved in being epistemically responsible (V). In Section 3.2, I consider and reject the first possibility. In Section 3.3, I briefly characterize some of the ways in which the IRS results will transform our epistemic practices, and I speculate about what sort of notion of epistemic responsibility can stand up to the IRS results. I suggest as an obvious possibility that in a post-IRS world, being epistemically responsible may well involve nothing more than using reliable reasoning strategies.

3.1. *The IRS Dilemma*

Do the IRS results raise problems for the internalist's understanding of epistemic responsibility? One might reasonably think they don't. Here's how someone might try to handle the IRS results concerning the admissions problem: "There are actually three different possibilities, and the internalist conception of epistemic responsibility can handle all three. (1) The cognitive agent is non-culpably ignorant of the IRS result. In this case, we cannot fault the agent for using the conventional strategy rather than the unit weight strategy. After all, epistemic responsibility does not require us to know anything about which we are non-culpably ignorant.[1] (2) The cognitive agent is culpably ignorant of the IRS results. In this case, if the agent does not use the IRS, we can reasonably judge her reasoning to be epistemically irresponsible. (3) Finally, suppose the cognitive agent knows the IRS results. In this case, it would be perfectly responsible for her to use a unit weight model to solve the admissions problem. *Given the IRS results, the unit weight prediction will satisfy the internalist virtue (or virtues) that constitute what it is to be epistemically responsible*. So as long as we're clear about the cognizer's epistemic relation to the IRS results, the internalist conception of epistemic responsibility is in no trouble." The main problem with this response is that the italicized sentence is false. It radically underestimates the chaos the IRS results will bring to our epistemic judgments. Or so I shall argue.

Perhaps the central task of epistemology is to provide some account for how to properly assess particular instances of reasoning. It is very easy to *say* that anyone who uses an IRS is being epistemically responsible. What's extremely difficult is to follow through on this when making judgments about particular instances of reasoning. The best way to appreciate the mayhem that the IRS results will bring to our judgments about what constitutes responsible reckoning is with an example. Consider two people who are aware of the IRS results. Beavis and Stuart are serving on an admissions committee and are at loggerheads over whether Smith or Jones will be the better candidate. Beavis argues, “When I added up their high school ranks and test score ranks, Smith’s score was 6 points higher than Jones’. Given the IRS results, we should go with Smith.”

“Well, I disagree,” responds Stuart. “Jones went to a more selective high school, which explains the lower class rank. Jones also took tougher courses. Further, some of Smith’s letters are a bit lukewarm. Jones’s letters are uniformly terrific.”

Beavis fumbles inefficiently through the file, nodding his head slowly. “Yeah, I guess that’s right. But I still think we should accept Smith. The IRS results show that in the long run, we’ll do better by just adding class ranks to test score ranks and admitting the candidates with the highest scores.”

“But, Beavis, don’t you think we should discount Smith’s higher GPA, since he took easier courses and went to a less selective high school than Jones? And don’t you think that letters of recommendation tell us something about the candidate’s abilities?”

Beavis ponders for a moment and replies, “Well, if we didn’t have the test scores and the high school rank, that evidence would be relevant, for sure. Using it would be more reliable than making random guesses. But since we do have the test scores and the high school rank, we’ll be more reliable in the long run if we focus only on them and weigh them equally.”

“I’ll grant that in the long run, your strategy is more reliable,” says Stuart. “But I want to know about *this* case. Doesn’t reason demand that in *this* case we should use our judgment, and not rely on a crude and mindless algorithm that ignores factors that you admit are relevant to predicting academic achievement? When the simple algorithm isn’t so obviously wrong, it’s a good idea to use it. But in this case, the algorithm makes clearly the wrong call. Overruling it is the epistemically responsible thing to do.”

Beavis replies quietly but firmly, “But in these situations, when people overrule the ‘mindless’ algorithm, more often than not, they’re wrong and the algorithm is right.”

Stuart replies, just as firmly, “But there will be cases in which the algorithm is wrong and human judgment is right. And we have lots of evidence for believing that this is one of those cases. If the algorithm makes a prediction that is clearly epistemically irresponsible, then epistemic responsibility demands that we reverse it.”

Seeing little support around the room, Beavis’s face darkens. “This whole conversation is starting to bum me out. Could we get on with it? I need to take a nap.”

While the story is fantasy, the stubborn refusal to implement IRS’s is very common, even among those who know the IRS results. Robyn Dawes recounts many examples of intelligent people who offer terrible arguments for ignoring the IRS results. The common feature of all these arguments is that they appeal to particular examples in which the IRS is deemed (rightly or wrongly) to be wrong or relatively unreliable. For example, Dawes tells about the reaction among clinicians at the Ann Arbor VA Hospital where he had implemented an IRS (the Goldberg formula from Section 1.2 and below).

> 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 particularly vivid example in which "I was talking about one study predicting death . . . [and] 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 . . . " (this volume; also Dawes 1994, 92). One detects a deep frustration in Dawes' writings with the robust resilience of this kind of response to IRS's, especially in the face of more than 40 years worth of completely one-sided evidence. But I want to suggest that the explanation for the powerfully resilient anti-IRS response is *not* that people are irrational or lack a basic commitment to epistemic responsibility. In fact, just the opposite. *It's just that we are committed to the wrong view of epistemic responsibility*. Or so I shall argue in Sections 3.2 and 3.3.

For now, I want to argue that it is Stuart, not Beavis, who best displays the virtue (or virtues) the epistemic internalist takes to be central to warrant or justification. To see this, consider two different strategies for solving the admissions problem: Beavis's unit weight model and the proper model. The internalist is surely committed to the view that the proper model's predictions best satisfy the virtue (or virtues) the internalist takes to be central to epistemic responsibility. After all, the proper model's predictions result from considering all the different lines of relevant, available evidence and properly weighing each line of evidence according to its predictive value. The proper model's predictions presumably optimize the belief-system's coherence, best fit the available evidence, have the best reasons in their favor, etc. Now, given the flat maximum principle, we know that for problems of social prediction, the unit weight model is about as reliable as the proper model. That does not mean that they always make the same predictions. What it does mean is that when they make different predictions, neither model is much more reliable than the other. It follows that the unit weight model will often make predictions that do not best satisfy the virtue (or virtues) internalists take to be central to epistemic responsibility. Now return to the disagreement between Stuart and Beavis. Stuart will always have an argument in favor of his prediction and against Beavis's that appeals to (a) evidence that Beavis has intentionally ignored, (b) the relative predictive powers of the cues, which Beavis has intentionally ignored, or (c) both. In many cases where Stuart and Beavis disagree, Stuart's prediction will be identical to the proper model's prediction. When this occurs, Stuart, and not Beavis, really does best satisfy the internalist virtue (or virtues) taken to be central to warrant or responsibility. *From the perspective of (V), Stuart is the responsible one. Beavis ignores evidence, does not calibrate the evidence he does consider in terms of its predictive power, and he makes predictions that controvert the available evidence.*

The fact remains, maddening though it may be: Beavis's predictions are more reliable than Stuart's. In fact, no strategy is more reliable than Beavis's. And no tractable strategy is as reliable. It is probably worthwhile to harp on this point. Faced with a problem of social prediction, the most reliable human agent will *not* attend to all relevant, available evidence. She will *not* weigh each piece of evidence in accordance with its predictive value. And this is not because of the reasonable and familiar cry that it is inappropriate to hold people to impossible epistemic standards. Ignoring most of the evidence *is* the most reliable way (or among the most reliable ways) to make social predictions, full stop. Not weighing each piece of evidence according to its predictive value *is* the most reliable way (or among the most reliable ways) to make social predictions, full stop. Further, reasoning in this manner is *more* reliable than Stuart's conventional labor-intensive strategy. It doesn't matter whether time and resources are at a premium or whether one has available eons, a supercomputer and a gaggle of mathematicians.

So here's the IRS dilemma in a nutshell. From the perspective of (V), Stuart is more responsible than Beavis. From the perspective of (C), Beavis is more responsible than Stuart.[2] So who is really more responsible?

We began this Section (3.1) with a very natural attempt to downplay the IRS results. The argument depended on the claim that when the IRS results are known, Beavis's predictions best satisfy the internalist virtue or virtues that are typically thought to be central to epistemic responsibility. But this can't be right. Beavis's predictions often do not satisfy the internalist's relevant epistemic virtue (or virtues). And yet, there is no more reliable way to make social predictions than to employ a unit weight model. So we face the IRS dilemma: When we restrict ourselves to the tractable strategies, to the strategies that normal, intelligent humans are capable of using when faced with a problem of social prediction, we can adopt the strategy that is most reliable or we can try our level best to make predictions that have the best reasons in their favor, that best "fit" or "cohere" with the evidence. But we can't do both. How should we reason?

One way out of the dilemma is to avoid it. The internalist might argue: "For problems of social prediction, we already know what the epistemically responsible agent will do – make the prediction sanctioned by the *proper* linear model. This view avoids the IRS dilemma, since the proper model satisfies (V) and is as reliable as the unit weight model. So I can keep (V) and consilience is not threatened." We can grant the internalist (and the externalist for that matter) that the maximally responsible cognizer will employ a proper linear model to make social predictions. But surely on anyone's view, a useful epistemology will spell out how us less-than-maximally responsible humans are supposed to reason. Kornblith notes, "Rules of ideal reasoning are not easily come by, and it does not always show some shortcoming on the part of the subject that his reasoning was less than ideal. Sometimes we wish to know whether a subject was reasoning 'as best he could' " (Kornblith 1983, 33). The IRS dilemma does not face a superhuman reasoner who is capable of constructing and solving proper linear models or Bayesian algorithms on the fly. But for real human reasoners, facing real problems of social prediction, what's the epistemically responsible way to reason? The irony is that if you, the reader, are inclined to avoid the IRS dilemma by appealing to a unit weight model, you will probably face a gatekeeping decision some day. And you won't have a proper linear model to help make that decision. You won't be able to claim non-culpable ignorance of the IRS results. (Too late now, you should have stopped reading earlier.) So will you reason responsibly? And what will that involve?

3.2. *Relinquishing Consilience*

So we're faced with a gatekeeping decision and the unit weight model makes a prediction that seems to clearly controvert the evidence. What do we do? We might go along with Stuart and so respect (V), the internalist conception of what it is to be epistemically responsible. But then we give up consilience, since Beavis's reasoning strategy is more reliable. Or we might go along with Beavis and retain consilience. But in that case, what of the epistemic virtues that we must display to be epistemically responsible? Are we to smother the siren call of (our conception of) epistemic responsibility as powerful and beguiling but ultimately illusory? I say *yes*. But let's first consider the other option.

Given what many philosophers say about epistemic responsibility, relinquishing consilience is a non-starter. According to Kornblith, an epistemically responsible agent "desires to have true beliefs" and acts on those desires (34), according to BonJour, she "accepts all and only those beliefs which [she] has good reason to think are true" (8), and according to Feldman and Conee, she "use[s her] evidence in trying to believe all and only the truths [she] considers" (20). If these descriptions are even roughly accurate depictions of (part of) what is involved in being epistemically responsible, then the epistemically responsible cognitive agent who knows the IRS results will go with Beavis.

My choice of passages about epistemic responsibility is biased, of course. Others might have brandished passages suggesting the notion of epistemic responsibility is conceptually tied to some virtue that Stuart displays and Beavis does not (or that Stuart displays to a greater degree than Beavis). While this kind of move is certainly open, it strikes me as thoroughly unmotivated. Given a choice between two strategies, where one is easier to use *and* more reliable, to insist that the laborious, less-reliable strategy is the epistemically responsible one threatens to empty epistemic responsibility of its action-guiding role. Unless the laborious strategy has some unseen benefit, surely the instrumentally rational person will choose the easier, more reliable strategy. And why not? Well, perhaps the answer is that embedded deep in analytic epistemology, in our epistemological practices and concepts, is a kind of earnest work ethic: Epistemic value increases with epistemic effort. If so, then so much the worse for our conventional epistemic practices and concepts. To willingly exert efforts that we know in the long run will bring more false beliefs is epistemological masochism. The IRS results should bring joy – they suggest that epistemic value increases with the occasional and well-timed epistemological siesta.

3.3. *Rethinking Epistemic Responsibility: Is Consilience all There Is?*

In the face of the IRS dilemma, the proper course is to retain consilience and assert that Beavis's reasoning is more responsible than Stuart's. What implications do the IRS results have for our epistemological practices? I will suggest three. First, and most obviously, the IRS results will force us to dramatically change the way we judge certain instances of reasoning. Second, we will be forced to endorse some very counterintuitive general epistemic prescriptions. And finally, the IRS dilemma will force us to rethink what it is for a cognitive agent to be epistemically responsible. I will suggest that that the IRS results make more plausible a generally reliabilist view of epistemic responsibility. To a very rough first approximation, to be epistemically responsible is to employ reliable belief-forming procedures. And insofar as epistemic responsibility is tied to justification, this would make more plausible a reliabilism about justification (the *locus classicus* is Goldman 1986).

3.3.1. *Epistemic Judgments Revised*

There is nothing revolutionary in the idea that we sometimes learn new and better ways to learn. To take a well-trod example, it was only after the placebo effect was discovered that we learned that double-blind experiments were important. The need for double-blind experiments did not spring a priori from our epistemological concepts. Like the IRS results, the placebo effect demanded a shift in our judgments about which instances of reasoning count as responsible and which do not. After the placebo effect, in making predictions about the effects of drug or treatments, responsible cognizers preferred the conclusions derived from double-blind experiments to those derived from single-blind experiments. After the IRS results, in making social predictions, responsible cognizers prefer conclusions derived from well-constructed unit weight models to those of human experts. But as I have tried to suggest, this analogy radically underestimates the havoc the IRS results will bring to our epistemic practices. In the Beavis-Stuart debate, Stuart's prediction better fits the evidence, and Stuart's belief-system is the more coherent. And yet, unless we are to give up consilience, we must judge Beavis to be the more responsible reasoner. For the sake of consilience, we must smother our inclination to attend to good reasons, to attend to fit with evidence, to attend to coherence. This is a nontrivial task. For example, there is a half century worth of experiments that show that unstructured interviews degrade human prediction (e.g., Bloom and Brundage 1947; DeVaul et al. 1957; Oskamp 1965; Mustein, et al. 1981), and yet most academic departments still conduct short, unstructured interviews when hiring. Even when we know the evidence, it's hard for us to believe that interviews are worse than irrelevant. We seem to get so much useful information from them! And yet, it is almost surely epistemically irresponsible to use unstructured interviews to make gatekeeping predictions. If it's fiercely difficult for us to relinquish just one kind of evidence (unstructured interviews), imagine how difficult it will be for us to relinquish (for specific kinds of reasoning problems) deeply held epistemic values, e.g., good reasons, fit with evidence, coherence, etc.

3.3.2. *Epistemic Prescriptions Revised*

When we revise our judgments about what instances of reasoning count as epistemically responsible, we need to bring those judgments into line with our general epistemic prescriptions. On Goodman's influential reflective equilibrium view, this process – or something like it – is all there is to justification: "The process of justification is the delicate one of making mutual adjustments between rules and accepted inferences; and in the agreement achieved lies the only justification needed for either" (1965, 64). For our purposes, we needn't accept the view that reflective equilibrium is all there is to the epistemological project. But it does seem plausible to suppose that part of that project involves bringing our judgments about particular instances of reasoning into line with our general epistemic prescriptions and *vice versa*. When we judge Beavis to be more epistemically responsible than Stuart, we are forced to make some rather shocking adjustments to the general epistemic prescriptions we are willing to endorse. In the post-IRS world, given a problem of social prediction, a responsible agent will *not* attend to all relevant, available evidence. She will *not* weigh each piece of evidence in accordance with its predictive value. In fact, if we restrict ourselves to tractable strategies, strategies that unaided humans of normal intelligence can employ, ignoring most of the evidence *is* the most responsible way to make social predictions, full stop. Not weighing each piece of evidence one does consider according to its predictive value *is* the most responsible way to make social predictions, full stop. Ignoring evidence and not calibrating the considered evidence is *more* responsible than the conventional, labor-intensive strategy of trying to make predictions that best fit the available, relevant evidence.

What's new and disturbing about the IRS results, what distinguishes them from the placebo effect, is that they bring in their wake epistemic judgments and prescriptions that challenge our most cherished views about the virtues involved in being epistemically responsible. For example, we accept the general prescription that certain kinds of study require double blind experiments, but that doesn't make us question whether the epistemically responsible agent must make judgments that best "fit" her evidence. On the other hand, the IRS results *do* make us question this. The prescriptions we are driven to endorse require us to often accept predictions that *don't* best fit our evidence. The IRS results tell us to believe a social prediction even in the face of powerful, undefeated, evidence-based reasons to believe a different, incompatible prediction.

3.3.3. *Epistemic Concepts Revised*

The IRS results suggest that philosophers have been much too cavalier in our faith that our current views about epistemic responsibility satisfy the consilience assumption. Given any particular conception of what epistemic responsibility involves, consilience is always a defeasible assumption. It is always possible that displaying some set of epistemic virtues is not the best way, or even among the better ways, to get at the truth. In fact, insofar as our current views about what epistemic responsibility involves push us to endorse Stuart's reasoning rather than Beavis's, they need to be amended. It's easy enough to *say* that it is epistemically responsible for Beavis to use the unit weight model. The real challenge is to integrate this IRS finding (as well as others) into our epistemic judgments about particular episodes of reasoning and into our general prescriptions about what counts as responsible reasoning.

When we do this, what is likely to happen to (V)? Right now, we have some idea of what it is for a person to be epistemically responsible. But those views need to be quashed in many cases. My own suspicion is that the IRS results will drive us toward a generally reliabilist view of epistemic responsibility, especially if more IRS's are discovered; that is, the more we are forced to quash (V) in our general epistemic prescriptions and in our judgments of particular episodes, the more likely we are to give up (V) as part of our conception of what it is to be epistemically responsible. If this happens, epistemic responsibility would be exhausted by consilience. To a very rough first approximation, being epistemically responsible would involve nothing other than employing reliable belief-forming procedures. Insofar as responsibility is tied to justification, the IRS results would also tend to make some kind of reliabilism about justification more attractive. Of course, filling out these sketchy suggestions in a plausible way is a big project better left for another day.

Finally, one might suspect that my predictions about the violence that is to come to our epistemic judgments, prescriptions and concepts are exaggerated. One might doubt that people will ever really embrace the IRS results, especially if it is so difficult for us to implement them. But the fact is, the transformation has already begun. While philosophers have, by and large, ignored the IRS results, many institutions that make predictions that have real and serious consequences for people's lives now use IRS's. When you apply for credit and when newborns are identified as at risk for Sudden Infant Death syndrome, it's almost surely an IRS, and probably a unit weight model, that's doing the predictive work (Lovie and Lovie 1986). It's not just that IRS's are coming. They're here, and they're here to stay.[3]

NOTES

[1] An interesting issue for another day is: Suppose the cognitive agent is non-culpably ignorant of the IRS result and the unit weight strategy is her default reasoning strategy. Would she be more responsible than if she'd used the conventional strategy? If Gigerenzer and the ABC group find fast and frugal heuristics that we are naturally inclined to use and that are as reliable as "ideal" reasoning strategies, this will not be a merely hypothetical issue.

[2] Some might be having some very powerful anti-IRS reactions. "Surely," one might think, "Stuart is right to overturn the IRS's prediction when he has good reasons to think the proper model's prediction would be different." Remember, given the flat maximum principle, for problems of social prediction, when the predictions of the proper model and the unit weight model disagree, neither is likely to be more probable. But there is empirical evidence as well to think Stuart's overrulings will more likely than not be mistaken. Goldberg (1968) trained subjects to make diagnoses of psychosis or neurosis on the basis of MMPI profiles by giving them immediate feedback on their predictions. He gave a simple formula (an IRS) to one group and told them that it would increase the reliability of their judgments. There was a short-term increase in accuracy, but "this effect gradually wore away over time" until it disappeared altogether (Goldberg 1968, 493). This suggests that when we know how the IRS works, when we know how crude and mindless it is, we are strongly tempted to be "epistemically responsible" and make end-runs around it. And by being "epistemically responsible" we degrade our reliability. Goldberg gave another group "the numerical value of the formula for each profile and the optimum cutting score." They were told that this formula would achieve approximately 70% accuracy and that it would be more accurate for extreme values than for values close to the cutting score." While this group "increased their accuracy to a bit below 70% correct ... the accuracy of these judges' diagnoses was not as high as would have been achieved by simply using the formula itself" (Goldberg 1968, 493). So even when the subjects only knew the IRS's reliability and not how simple, crude and mindless it was, Goldberg's study suggests that we still can't improve on the IRS. When we overrule the IRS for the sake of epistemic responsibility, our overrulings are wrong more often than not. Our commitment to "epistemic responsibility" leads us to make unreliable – but typically very confident – judgments.

[3] Thanks to Richard Samuels for very helpful comments on an earlier draft of this paper.

REFERENCES

Bloom, R. E. and E. G. Brundage: 1947, 'Predictions of Success in Elementary School for Enlisted Personnel', in D. B. Stuit (ed.), *Personnel Research and Test Development in the Naval Bureau of Personnel*, Princeton University Press, Princeton, pp. 233–61.

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

Bowman, E. H.: 1963, 'Consistency and Optimality in Managerial Decision Making', *Management Science* **9**, 310–21.

Cohen, L. J.: 1981, 'Can Human Irrationality be Experimentally Demonstrated', *Behavioral and Brain Sciences* **4**.

Cohen, S.: 1983, 'Justification and Truth', *Philosophical Studies* **46**, 279–95.

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 Beauty of Improper Linear Models in Decision Making', in D. Kahneman, P. Slovic and A. Tversky (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, Macmillan, New York.

Dawes, R.: 1999, 'A Theory of Rationality as a 'Reasonable' Response to an Incomplete Specification', this issue.

Dawes, R. and B. Corrigan: 1974, 'Linear Models in Decision Making', *Psychological Bulletin* **81**, 95–106.

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

Einhorn, H. J. and R. M. Hogarth: 1975, 'Unit Weighting Schemas for Decision Making', *Organizational Behavior and Human Performance* **13**, 172–92.

Feldman, R. and E. Conee: 1985, 'Evidentialism', *Philosophical Studies* **48**, 15–34.

Gigerenzer, G., J. Czerlinski, and L. Martignon: 1999a, 'How Good are Fast and Frugal Heuristics?', in J. Shanteau, B. Mellers, and D. Schum (eds), *Decision Science and Technology: Reflections on the Contributions of Ward Edwards*, Kluwer, Norwell, MA.

Gigerenzer, G. and D. G. Goldstein: 1996, 'Reasoning the Fast and Frugal Way: Models of Bounded Rationality', *Psychological Review* **98**, 506–26.

Gigerenzer, G., P. Todd and the ABC group (eds.): 1999, *Simple Heuristics that Make us Smart*, Oxford University Press, New York.

Gilovich, T.: 1991, *How We Know What Isn't So*, The Free Press, Macmillan, New York.

Goldberg, L.: 1968, 'Simple Models of Simple Processes? Some Research on Clinical Judgments', *American Psychologist* **23**, 483–96.

Goldberg, L.: 1970, 'Man vs. Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences', *Psychological Bulletin* **73**, 422–32.

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

Goldstein, D. and G. Gigerenzer: 1999, 'The Recognition Heuristic: How Ignorance Makes us Smart', in G. Gigerenzer, P. Todd, and the ABC group (eds): In Press, *Simple Heuristics that Make Us Smart*, Oxford University Press, New York.

Goodman, H.: 1965, *Fact, Fiction and Forecast*, Bobbs-Merrill, Indianapolis and New York.

Kahneman, D. and A. Tversky: 1973/82, 'On the Psychology of Prediction', *Psychological Review* **80**, 237–51. Reprinted in D. Kahneman, P. Slovic, and A. Tversky (eds), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge, pp. 48–68.

Kornblith, H.: 1983, 'Justified Belief and Epistemically Responsible Action', *The Philosophical Review* **92**, 33–48.

Lovie, A. D. and P. Lovie: 1986, 'The Flat Maximum Effect and Linear Scoring Models for Prediction', *Journal of Forecasting* **5**, 159–68.

Martignon, L., U. Hoffrage, and N. Kriegeskorte: Manuscript, 'Lexicographic Comparison Under Uncertainty: A Satisficing Cognitive Algorithm'.

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

Milstein, R. M., L. Wilkinson, G. N. Burrow, and W. Kessen: 1981, 'Admission Decisions and Performance During Medical School', *Journal of Medical Education* **56**, 77–82.

Nisbett, R. and L. Ross: 1980, *Human Inference: Strategies and Shortcomings of Social Judgment*, Prentice-Hall, Englewood Cliffs, NJ.

Oskamp, S.: 1965, 'Overconfidence in Case Study Judgments', *Journal of Consulting Psychology* **63**, 81–97.

Payne, J. W., J. R. Bettman, and E. J. Johnson: 1988, 'Adaptive Strategy Selection in Decision Making', *Journal of Experimental Psychology: Learning Memory, and Cognition* **14**, 534–52.

Payne, J. W., J. R. Bettman, and E. J. Johnson: 1993, *The Adaptive Decision Maker*, Cambridge University Press, New York.

Plantinga, Alvin: 1993, *Warrant: The Current Debate*, Oxford University Press, New York.

Tversky, A., and D. Kahneman: 1974, 'Judgment Under Uncertainty: Heuristics and Biases', *Science* **185**, 1124–31.

Wiggins, N. and E. S. Kohen: 1971, 'Man vs. Model of Man Revisited: The Forecasting of Graduate School Success', *Journal of Personality and Social Psychology* **19**, 100–106.

Department of Philosophy
Iowa State University
402 Catt Hall
Ames IA 50011
U.S.A.