11

ALVIN GOLDMAN

Human Rationality: Epistemological and Psychological Perspectives

Both philosophers and psychologists have had extended debates over the prospects for human rationality. In recent years these debates have included interchanges across the disciplines, and reciprocal influences on one another. They provide a perfect context for exploring 'naturalizing' projects in philosophy, especially in epistemology. Four questions can be asked here: Do humans have the cognitive wherewithal to be rational? (Epistemic rationality is our focus.) What criterion of rationality should be employed in assessing these prospects? Which disciplines should tackle these questions: philosophy, empirical psychology, or a combination of the two? If these issues require a collaborative enterprise between philosophy and psychology, how exactly should this collaborative enterprise be structured?

My discussion is initially organized around two approaches to rationality: (1) the *formal-rules* approach and (2) the *good outcomes* approach. This organizing principle is partly guided by the fact that philosophical theories of normativity are usually divided into consequentialist and non-consequentialist varieties. It is also motivated by the fact that psychologists who debate rationality commonly appeal either to rule-based or outcomes-based criteria of rationality.

1 The Formal Rules Approach and the Psychological Evidence

The psychological debate over human rationality originates in the 'heuristics and biases' program pioneered by Amos Tversky and

Daniel Kahneman (Kahneman *et al.* 1982). According to this program, experimental studies of human reasoning show that people do not reason in accordance with normatively correct principles or rules of reasoning. Instead, they employ psychologically simpler principles that sometimes get right answers and sometimes do not. At least five types of tasks or outcomes allegedly reveal erroneous, biased, or counter-normative judgments or choices: (1) the selection task, (2) the conjunction fallacy, (3) base-rate neglect, (4) overconfidence, and (5) framing effects. I shall briefly explain three of them, each of which has been extensively replicated.

In the original selection task (Wason 1966), subjects were shown four cards, two with a number on the facing side and two with a letter. They were told that all cards have a number on one side and a letter on the other. They were then asked to identify the cards that must be turned over to determine whether the following rule holds: 'If a card has a vowel on one side, then it has an odd number on the other?' This is a question of logic, to which the correct answer is that two specific cards must be turned over. Typically, fewer than 10% of (naïve) subjects name exactly the right cards.

In another type of experiment demonstrating the 'conjunction effect,' subjects were given a description of an individual, say Linda, and asked to rank the probabilities of a list of statements about her. For example, (A) Linda is a bank teller, or (B) Linda is a bank teller and active in the feminist movement. Probability theory entails that the probability of a conjunction can never exceed the probability of one of its conjuncts. But about 89% of naïve subjects rank the probability of (B) higher than the probability of (A) (Tversky and Kahneman 1983).

In a third type of study, subjects were told that thumbnail descriptions had been constructed of 70 engineers and 30 lawyers (or 30 engineers and 70 lawyers). They were given five of these descriptions and asked to indicate the probability that the individual is an engineer or a lawyer. A proper estimate should reflect the base-rate of the specified population, i.e., either a 70/30 ratio or a 30/70 ratio (depending on what they were told). In fact, most subjects ignored the base-rate (Kahneman and Tversky 1973).

These and similar propensities to violate rules of logic or probability led proponents of the heuristics and biases program (H&Bers) to speak of "bleak implications for human rationality" (Nisbett and

Borgida 1975). H&Bers not only expose (alleged) violations of epistemic norms but propose cognitive heuristics responsible for these violations, such as the 'representativeness' heuristic (Tversky and Kahneman 1974). However, assorted critics, both philosophers and psychologists, have offered a variety of responses that preach greater optimism about human rationality. Here is a brief overview of such responses that appear in the literature, by both philosophers and psychologists.

Philosophers' reject-the-norm responses. Philosophers L. Jonathan Cohen (1981) and Daniel Dennett (1987) offer general approaches to rationality according to which it is *conceptually* impossible for people to be systematically irrational. People can make 'performance errors' — e.g., computational mistakes or memory lapses — but they cannot be systematically irrational. This is because reasoning 'competence' is constituted by human dispositions to reason. To paraphrase Protagoras, man (i.e., human nature) is the measure of rationality.

Psychologists' reject-the-norm-responses. The psychologist Gerd Gigerenzer (1991a) questions the interpretation of probability presupposed by many H&B studies. The Linda task and the engineer-lawyer task assume probabilities for single cases. But it is controversial whether single-case probabilities make sense. A purely frequentist interpretation of probability denies the meaningfulness of such statements. So there are no norms here for subjects to violate. This is one instance of a general form of response to alleged norm violations: deny that the invoked norm is a legitimate or *bona fide* norm.

Resource Relativization. According to some approaches (Simon 1957; Goldman 1978, 1993; Cherniak 1986; Stich 1990; Harman 1995; Osherson 1995; Gigerenzer and Goldstein 1996), rationality evaluations should be relativized to the resources available to the human cognitive apparatus. If the human apparatus does not include competence vis-à-vis certain tasks, then failure to perform 'correctly' on those tasks is not a case of irrationality. Stanovich (1999) expresses this idea by distinguishing between 'normative' and 'prescriptive' questions. Normative questions concern what is required by formal rules. Prescriptive questions concern what is best for a reasoner to do given an admittedly limited cognitive repertoire.

Finding reasoning competences. An implication drawn by H&Bers

from their studies is that humans lack a correct grasp of either logic or probability. In response, so-called 'evolutionary psychologists' such as Gigerenzer, Cosmides and Tooby (Gigerenzer and Hoffrage 1995; Gigerenzer 1991a, 1998; Cosmides and Tooby 1992, 1996) concede that people do not have mental formats for representing logical relations in their full, content-neutral generality, and do not have mental formats for representing probabilities in terms of percentages. However, these psychologists argue that people do have evolved mental modules for dealing correctly with special problems involving logic and probability that the species encountered in its evolutionary history. For example, a special module enables people to deal correctly with selection tasks that concern the maintenance of social exchange relationships. And probabilities can be represented and understood as 'natural frequencies' learned by 'natural sampling.'

Starting with the probability case, Cosmides and Tooby write: "Our hominid ancestors were immersed in a rich flow of observable frequencies that could be used to improve decision-making, given procedures that could take advantage of them. So if we have adaptations for inductive reasoning, they should take frequency information as input" (Cosmides and Tooby 1996, pp. 15–16). As Gigerenzer puts it, this information was encoded "as frequencies of events, sequentially encoded as experienced — for example, *3 out of 20* as opposed to 15% or p = 0.15" (Gigerenzer 1994, p. 142).

This speculation led Cosmides and Tooby to perform a series of experiments in which they transformed the "Harvard Medical School problem" of Casscells *et al.* (1978). In the original study, Casscells *et al.* had physicians and medical students answer questions about, for example, the likelihood of a patient's having breast cancer given information about the case that was formulated in a probability format. Only eighteen percent of the subjects gave the correct Bayesian answer. Cosmides and Tooby (1996) did new experiments in which subjects were given essentially the same problem but presented in frequency terms. For example, subjects were told, 'Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease.' In this alternative formulation, 76 percent of the subjects gave a correct Bayesian response. Further experiments were designed to explore differences between methods of formulation. The two factors with the strongest effect on accuracy of performance, Cosmides and Tooby reported, involved use of a frequency format:

"Asking for the answer as a frequency produces the largest effect, followed closely by presenting the problem information as frequencies" (Cosmides and Tooby 1996, p. 58).

Another investigator, Fiedler (1988), produced similar evidence. Following up on evidence presented by Tversky and Kahneman themselves (1983), the Linda problem was presented to subjects with frequentist wording: "There are 100 people who fit the description above. How many of them are: ... bank tellers? ... bank tellers and active in the feminist movement?" (Fiedler 1988, p. 25). Only 22% of Fiedler's subjects judged that there would be more feminist bank tellers than bank tellers. Similarly, Hertwig and Gigerenzer (1994), reported in Gigerenzer (1994), told subjects that there were two hundred women fitting the 'Linda' description and asked them to estimate the number who were bank tellers, feminist bank tellers, and feminists. Only 13 percent committed the conjunction fallacy.

Next consider studies by evolutionary psychologists concerning Wason's selection task. Early investigators noticed that although subjects perform poorly on most versions of the selection task, their performance on some versions improves dramatically. Griggs and Cox (1982) gave subjects a form of the task in which the rule was, 'If a person is drinking beer, then he must be over 20 years old.' They were shown four cards, each representing one person. One side of the card said what the person is drinking and the other side gave the person's age. Subjects were asked: exactly which cards must be turned over to see if any of these people were breaking the law? About 75 percent of the subjects made the correct two-card selection, a far better performance than in other versions of the task. To explain this and other so-called 'content effects,' Cosmides and Tooby (1992) speculated that if stable social exchange arrangements are to exist in a society, its members must have cognitive mechanisms that enable them to detect 'cheaters' who defect from these arrangements for their own advantage. Evolutionary analysis led them to expect that there must be one or more mental modules dedicated to detecting cheaters — people who accept the benefits of social contracts or social arrangements but do not pay the costs. Griggs and Cox's beer-drinking rule is an example of a selection task that would trigger a cheater detection module. Cosmides and Tooby assembled an impressive array of evidence in support of this hypothesis.

What conclusions should be drawn from this debate? As Samuels

and Stich (2004) point out, there are weaker and stronger ways to interpret the evidence. Evolutionary psychologists could summarize their evidence by saying that there are *some* reasoning problems on which people's intuitive judgments do not deviate from appropriate norms. Similarly, H&Bers could summarize their evidence by saying that there are *some* reasoning problems on which people's intuitive judgment do deviate from appropriate norms. On the other hand, many proponents on each side of the dispute state their respective conclusions much more strongly. Some adherents of the H&B approach claim that the only cognitive tools available to untutored people are normatively problematic, while some formulations of the evolutionary psychology critique seem to imply that human rationality problems are not very serious at all. As an example of the latter, Gigerenzer (1991a) wrote about "mak[ing] cognitive illusions disappear," and he has said things like, "We need not necessarily worry about human rationality," (Gigerenzer 1998, p. 280) and "intuition [is] basically rational" (Gigerenzer 1991b, p. 242). Samuels, Stich and Bishop defend a compromise position: "People do make serious and systematic errors on many reasoning tasks, but they also perform quite well on many others" (Samuels and Stich 2004, p. 296; Samuels *et al.* 2002).

2 What Is the Criterion for Cognitive Rationality?

Before deciding which of these positions to endorse, there is clearly a prior question that needs to be addressed. We cannot identify the extent to which human reasoning mechanisms violate or conform to norms of rationality unless the norms are antecedently specified. Where do we find a list of these norms? Perhaps we do not need a full enumeration of them, but we at least need a *criterion* of rationality such that we can tell of any proposed norm of rationality whether it really is one. There are two problems here. First, we need a way to determine which putative norms are genuine norms of rationality. Second, we need a formula, or set of formulas, for applying genuine norms to particular cases. Recall that Gigerenzer questioned whether probabilistic norms apply to single cases such as Linda's being a bank teller or Linda's being a feminist and a bank teller. This illustrates the problem of norm application. However, I shall focus on

the first and more fundamental problem: What is the criterion for (correct) norms of rationality?

As indicated earlier, the most common approach — especially among H&Bers — is to associate principles of rationality with formal rules that can be extracted from subjects like deductive logic, probability theory, and statistics. It is assumed that deductive logic, probability theory, and statistics provide standards of how the mind ought to operate. In other words, there are mappings from systems of logic or probability to cognitively or epistemically proper behavior. One problem here is that there is no unique system of logic, or probability, or statistics (although there are pretty standard versions of each, especially logic and probability). Which system of each kind is the appropriate one on which to base rules of rationality? I shall set this issue aside. For argument's sake, let us assume that there is a single formal system in each domain. That leaves open another serious question: Are there any simple mappings from the contents of these systems to appropriate norms of epistemic rationality?

One straightforward problem is that the subject-matters of the formal systems in question are not obviously identical to the subject-matter of epistemic norms. This is particularly clear in the case of logic. Logic is concerned with two kinds of propositions: propositions about semantic matters, such as satisfiability in a model and strict implication, and propositions about syntactic matters, such as provability in a system. None of these types of propositions has anything to do with the distinctive subject-matter of epistemology, namely, doxastic states such as belief, disbelief, or degrees of belief (subjective probabilities). Nor do they formulate *norms* for the adoption of doxastic states. It is quite unclear, therefore, how norms of rationality can literally be derived from propositions of logic.

Consider, however, some alleged norms of epistemic rationality that many writers somehow associate with the 'dictates' of logic. A commonly invoked rule of rationality is inconsistency avoidance: do not believe any logically inconsistent set of propositions. This rule of epistemic rationality, however, has been questioned by recent epistemologists (e.g., Foley 1987; Nozick 1993). Both in everyday life and in science we try to avoid inconsistencies. But when we find ourselves caught in an inconsistency, what does rationality dictate? Should we give up *everything* we believe? Surely not. Which beliefs, then, should be discarded? Very often this is not clear. Scientists

working with an inconsistent theory frequently bide their time and refuse to renounce all of its predictions. Thus, it seems to be rational to retain an inconsistent set of beliefs even when we know it is inconsistent. This cuts against a simple mapping between inconsistency and irrationality.

Let us think a moment longer about inconsistency avoidance. How bad is inconsistency, after all, from an epistemic perspective? Should inconsistency avoidance really be a cardinal principle of epistemic rationality? Holding an inconsistent set of beliefs has negative ramifications for truth-possession, and this is something with which rationality should be concerned. But just how bad are the ramifications? Holding an inconsistent set of beliefs only guarantees the falsity of at least one belief. It does not guarantee anything worse than this. On the other hand, if an agent abandons some of her beliefs to remove an inconsistency, she runs a distinct risk of losing some true beliefs previously possessed. So it is not obvious that a proper norm of rationality enjoins immediate abandonment of some beliefs (which ones?) when one finds oneself in an inconsistency. And no such norm follows from logic.

Now consider another possible norm of rationality associated with logic: deductive closure. This says that it is rational to infer all the logical consequences of whatever you believe. Again, this norm does not follow from logic *per se*. Logic says nothing about 'inference' in the sense of inferring a *belief*, because logic does not deal with belief or other doxastic states. Moreover, this principle cannot be followed, because there are infinitely many logical consequences of any proposition but people are incapable of believing infinitely many things. Even a weaker rule of permission allowing one to infer *any* deductive consequence of prior beliefs is epistemologically unacceptable. If R is deducible from my prior beliefs P and Q, but R conflicts with a current perceptual experience or is patently absurd (as judged by other beliefs), I should not automatically proceed to form a belief in R. Better to reject P or reject Q instead. So there is no obvious transition from logic to normative principles of cognition (Harman 1973, 1995; Goldman 1986; Nozick 1993).

Here are additional examples to show that there is no simple connection between truths of a formal system and rational inference. Suppose P is a random truth of a formal system, but not widely recognized. Suppose you have strong *testimonial* evidence of its falsity.

You hear a prominent expert in the field declare in a lecture that P is false (without offering a detailed proof). Is it *irrational* of you to reject P or withhold judgment about P rather than believe it? No. Rejecting P would be eminently reasonable, despite the fact that, by hypothesis, it is a formal (semantic or proof-theoretic) truth. Thus, a proposition's being a formal truth does not imply that a rational person, no matter what his circumstance, should believe it. Here is another example. Let P be a theorem of some suitable formal system, but one that has never been proved. Is it rationally mandatory for each person to believe P? Surely not.

A different possible link between formal truth and rational attitude is that if P is formally true *and* you possess good evidence for it, then it is rational for you to believe it. The trouble with this proposal is that only the second conjunct of the antecedent (the evidence possession conjunct) may be doing the work. In other words, it may be a correct principle of rationality that if you possess strong (and undefeated) evidence for *any* proposition, whether true or false, it is rational to believe it. This undercuts an intimate connection between *correctness* of formal rules and rational belief or inference. Furthermore, this rationality principle may not support the H&B-favored conclusions from the experimental literature, because it may be unclear that the subjects have evidence for the truths in question. During problem-solving moments, at any rate, they may not *actively* possess such evidence.

Let me express the latter point a bit differently. Our previous considerations suggest that failure to believe or deploy a relevant formal truth in problem solving is irrational when — but only when — the truth is evidentially *accessible* to the cognizer. When the truth has never been proved, or when you have negative testimonial evidence from an expert, the truth is not epistemically accessible. I now suggest that in problem-solving contexts, accessibility of evidence should be relativized to the search-space that the problem induces. A central task for any problem solver is to navigate the search space, to find information in one's head that is relevant to the task at hand. When the formulation or presentation of the problem naturally leads a problem solver to go down certain search paths rather than others, some potentially relevant information may be quite inaccessible. This information just won't be retrieved; it won't 'come to mind.' This may not be a defect in human problem-solving architecture. The hu-

man mind is remarkable for its ability to call relevant information to mind when confronting various tasks or problems, but it is not perfect. Does this mean that our cognitive architecture is simply irrational? I do not think so. It is not as if Artificial Intelligencers have designed computational engines that surpass the human mind in terms of the retrieval of relevant information.

These considerations bear on some of the previously cited H&B studies. In the original Linda study, for example, when subjects asked to rank the probabilities of Linda's possible vocations or avocations, it is not obvious to them that it is relevant to consider the conjunction theorem of probability. No doubt, the conjunction principle is a truth of the probability calculus; it may even be one that subjects have learned at some juncture in learning probability theory. But that does not entail that it will be epistemically accessible *given the task demands*. Similarly, subjects confronted with a lawyer/engineer problem might naturally focus their minds on information contained in the description of the target individual. The relevance of the engineer-lawyer ratio might not be transparent from this vantage-point. So although subjects in these experiments make errors of omission in their reasoning — 'omission' insofar as they neglect to apply relevant truths or principles — this does not establish their *irrationality*. (This response piggy-backs on the 'resource-relativization' proposal discussed in section 1.)

What are the morals of this discussion to the respective roles of philosophy and empirical psychology in the rationality debate? If rationality standards can be pinpointed, experimental psychology can then show that human cognitive endowments either meet those standards, fail to meet them, or meet them to some degree or other. But not all contributions to the rationality question can come from experimental psychology. In particular, specification of appropriate standards must come — in whole or part — from philosophy. H&Bers assume that norms or standards of rationality follow from truths of probability, statistics, and the like, which are well known to psychologists who use these tools in their science. However, although most psychologists have an excellent grasp of logic, probability, and statistics, this grasp is not the product of empirical psychology. Logic, probability, and statistics are not products (discoveries) of psychological science; they are materials *used* in psychological science. Moreover, as I have argued, there is no simple mapping from logic or

probability to rationality principles. So, which discipline is responsible for determining the correct principles of rationality? This belongs to philosophy, and in particular to epistemology.

3 The Consequentialist Approach

The consequentialist approach says that the rationality of a reasoning process is a function of the consequences of using it in a wide range of cases. Which types of consequences? In epistemology a clear candidate for the relevant kind of consequences is the proportion of true, or accurate, beliefs that the process-type produces. Thus, a reasoning process is rational if and only if it is *reliable*—i.e., usually generates true judgments or beliefs rather than false ones. Admittedly, reliability has rarely been proposed as a criterion of *rationality*. It is usually offered as a criterion of another normative epistemic status: *justification* (Goldman 1979, 1986). But many epistemologists treat justification and rationality as indistinguishable from one another, or at least intimately related. So reliabilism is one possible alternative to the formal-rules approach to rationality.

In fact, each of the other contributors to this symposium, Gerd Gigerenzer and Michael Bishop, has made roughly this proposal; each regards truth, or accuracy, as (at least) one type of consequence that is pertinent to a rationality appraisal. They do not always use the term 'rational' to identify the value of interest. But since they each propose criteria for good reasoning or inference strategies, this interpretation is not fanciful.

Gigerenzer writes about good inferential strategies being "fast and frugal" and still making one "smart" (Gigerenzer *et al.* 1999). This idea is further spelled out as follows:

> We propose a competition between various inferential strategies. The contest will go to the strategy that scores the highest proportion of correct inferences (accuracy) using the smallest number of cues (frugality). (Gigerenzer 2000, p. 177; based on Gigerenzer and Goldstein 1996)

So the goodness of an inferential strategy is equated with a propensity to produce high accuracy at low cognitive cost, i.e., with few cues.

Bishop and Trout (2005) make similar proposals, explicitly appealing to reliability in the epistemologist's sense. Their favored epistemological view is called *strategic reliabilism*, which they characterize as follows:

> Epistemic excellence involves the efficient allocation of cognitive resources to robustly reliable reasoning strategies applied to significant problems. (Bishop and Trout 2005, p. 71)

A 'robustly' reliable reasoning strategy is one that is reliable across a wide range of environments. Bishop and Trout use the normative phrase 'epistemic excellence,' and emphasize that they are interested in cognitive normativity in the sense that yields "explicit, reason-guiding advice on how people ought to reason" (Bishop and Trout 2005, p. 15). It is not clear that they equate epistemic excellence with rationality, but I shall make this interpretive assumption.

What do Bishop and Trout mean by a 'reasoning strategy'? Their idea of a reasoning strategy is illustrated by ten examples of *statistical prediction rules*, originally developed by the psychologist Paul Meehl (1954). An example of statistical prediction rules is the 'Goldberg Rule' (based on Goldberg [1965]). It predicts whether a psychiatric patient is neurotic or psychotic on the basis of a MMPI (Minnesota Multiphasic Personality Inventory) profile, where certain weights are given to the patient's scores on various clinical scales of the MMPI. Citing evidence that the Goldberg rule is a more reliable predictor of psychiatric traits than clinical experts, Bishop and Trout say that people *ought* to make use of the Goldberg Rule in making preliminary diagnoses of psychiatric patients (Bishop and Trout 2005, p. 15). Their conclusion about rationality seems to be: it is rational to use a given statistical prediction rule (SPR) only if it is robustly reliable. Thus, the prototype of a reasoning 'strategy' for Bishop and Trout is a formula or algorithm that specifies a set of cues and a mathematical weighting of these cues, on the basis of which outcomes can be predicted. When such a formula is robustly reliable, and is more reliable than any alternative rule for solving the same problems (Bishop and Trout 2005, pp. 76–77), it is rational to use this rule. Bishop and Trout add a number of further complications involving costs and benefits, but these can be set aside here.

Is this a plausible theory of epistemic rationality? No. Using a maximally reliable strategy (of those available) is neither necessary nor sufficient for rationality. It is not necessary because one can rationally use a strategy that is less reliable than another if one is nonculpably ignorant of the most reliable strategy. Using a maximally reliable strategy is not sufficient for rationality because the most reliable strategy might be used by dumb luck, in which case its use is not rational.

An SPR is a kind of algorithm that has been shown, by some sort of empirical research, to be highly reliable at predicting certain outcomes. The SPRs recommended by Bishop and Trout have been shown to be more accurate predictors than experts who use traditional methods. For example, Bishop and Trout (2005, p. 13) report that SPRs predict which newborns are at risk for Sudden Infant Death Syndrome (SIDS) much better than human experts do (Carpenter *et al.* 1977; Golding *et al.* 1985), and there is an SPR that more reliably predicts the quality of the vintage for a red Bordeaux wine decades in advance than do expert wine tasters, who swirl, smell, and taste the young wine (Ashenfelter *et al.* 1995). Does this mean that somebody who uses such SPRs to make appropriate predictions is rational, and somebody who relies on superseded experts is irrational? No. Someone might happen to adopt an optimal SPR without evidence of its optimality; and someone might bypass an optimal SPR in favor of experts because she is unaware of its demonstrated superiority. SPRs are like cooking recipes. You might choose the best recipe because your finger chanced upon it in a borrowed cookbook; that does not make you rational.

How can I, in all self-consistency, be so critical of a reliabilist criterion of epistemic normativity? Have I not myself defended reliabilism about epistemic normativity (Goldman 1979, 1986, and elsewhere)? How does my approach circumvent the very problems just identified for Bishop and Trout?

Several points are germane here. First, my brand of reliabilism has been applied only to justification and knowledge, not rationality. Second, the reliabilist theory defended in most of my work has used a non-comparative, satisficing criterion rather than a comparative, maximizing criterion (Goldman 1986, pp. 104–106). Third, the reliabilist theory defended in *Epistemology and Cognition* (Goldman 1986) included some crucial nuances that distinguish it from Bishop

and Trout's theory. Let me elaborate on those nuances that are especially pertinent here.

Epistemology and Cognition (pp. 92–95) distinguished between *processes* and *methods*. Processes are 'basic' or 'elementary' psychological operations (or sequences of such). Methods correspond to the strategies or rules discussed by Bishop and Trout. Methods are in no sense basic psychological operations. They are not basic because they need to be acquired by a learning process, often mediated by culture. They are not psychological in any interesting sense, although they can be the *contents* of psychological states and can guide behavior or conduct just like a cooking recipe. This does not make a recipe a psychological process in any interesting sense. (*Following* a cooking recipe is a psychological process. But indefinitely many psychological processes might realize the same recipe, so following a given method is not a psychological kind.)

As Bishop and Trout indicate, a predictive strategy can be instantiated by running a computer program that implements that strategy. This could be executed by simply clicking a button on your computer, thereby launching the program in question. A button might be clicked for a robustly reliable method while being totally ignorant of its reliabilist virtues. That would not be a specimen of rationality. In *Epistemology and Cognition* I addressed this kind of problem (directed at justifiedness rather than rationality) by distinguishing between 'primary' and 'secondary' levels of justifiedness. Primary justifiedness is mainly a matter of process-level reliability, and secondary justifiedness is partly a matter of method-level reliability. Full justifiedness is a function of both. What is important for present purposes is that it is not sufficient for secondary justifiedness that the agent use a rule or method that happens to be reliable. The agent must also select and acquire the method by a suitable *process*, in particular, by a process that is meta-reliable, i.e., tends to select reliable methods. Bishop and Trout's approach lacks this important wrinkle.

4 Psychology's Role in Naturalized Epistemology

That 'methods' or 'strategies' are not psychological entities is important to the issue of psychology's role in addressing rationality. Bishop and Trout argue for a radical version of psychologistic

epistemology, which tries to dispense with traditional epistemology entirely. 'Standard Analytic Epistemology,' they proclaim, is bankrupt. It should be replaced by what they call 'Ameliorative Psychology,' which would conduct the work of applied epistemology by identifying robustly reliable strategies, thereby promoting epistemic excellence. As they acknowledge, this would mean a total usurpation of epistemology by psychology, leaving nothing for Standard Analytic Epistemology to do.

In *Epistemology and Cognition*, I advocated a fairly substantial role for empirical psychology within epistemology, a thesis many philosophers regarded as fairly radical (in its day). Bishop and Trout mean to advance a still more radical vision of epistemology. Their conception of 'reformed epistemology' is apparently predicated on the assumption that identification of all good strategies depends on empirical psychology. They clearly imply that identification of robustly reliable strategies in general, and statistical prediction rules in particular, is (A) a core responsibility of applied epistemology, and (B) a job for psychology. This is exactly what 'Ameliorative Psychology' is intended to do.

This radical thesis strikes me as untenable. Notice first that strategies, rules, and recipes can have any content whatsoever. They can involve physical, chemical, geological, or economic phenomena. For any such domain, there is no end of possible rules for predicting or retrodicting events in the domain. Obviously, the job of determining a given rule's reliability should be assigned to the discipline — typically a scientific discipline — that studies the relevant domain. There is no reason to think, however, that psychology is the appropriate discipline for *all domains*. Bishop and Trout fall into the trap of thinking so because, historically, scientific psychologists wanted to expose the overblown claims to reliability made by their clinical brethren. These psychologists showed that there were better methods than the clinicians were using to make their predictions. However, it is a mistake to generalize from this local episode in the history of science to a privileged epistemological role for psychology. When it comes to what I call reliable *methods* — as opposed to *processes* — psychology has no special role to play. It is no more important to normative applied epistemology than any science that provides truths that are transformable into domain-specific strategies of inference.

However, I agree that psychology has a special role to play in epistemology, because it alone studies basic cognitive processes, i.e., representational and credence-forming capacities of the human mind (Goldman 1986). Psychology should provide detailed information about the characteristics of mental mechanisms, especially reasoning mechanisms. Then, whatever criterion or standard is antecedently identified as essential to rationality, it can be determined whether our mental mechanisms (or which of them) meet that standard or violate it.

It is the job of philosophical epistemology to determine the appropriate criterion or standard. Is it conformity to a certain class of formal rules? Is it a matter of tending to produce true or accurate judgments, or some other intellectual outcomes? Bishop and Trout presuppose, with little discussion, that rationality or epistemic excellence is fundamentally a matter of reliability. In advocating reliabilism, they share the same general epistemological orientation I do. What they do not appreciate, however, is that it takes philosophical argument to *show* (or try to show) that this is the appropriate standard. Thus, contrary to their view, there is a core, foundational role to be played in the enterprise by 'Standard Analytic Epistemology.'

REFERENCES

Ashenfelter, O., Ashmore, D. and Lalonde, R., 1995. Bordeaux wine vintage quality and the weather. *Chance*, 8, 7–14.

Bishop, M.A. and Trout, J.D., 2005. *Epistemology and the Psychology of Human Judgment*. Oxford: Oxford University Press.

Carpenter, R., Gardner, A., McWeeny, P. and Emery, J., 1977. Multistage score system for identifying infants at risk of unexpected death. *Archives of Diseases in Childhood*, 53, 606–612.

Casscells, W., Schoenberger, A., and Grayboys, T., 1978. Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, 299, 999–1000.

Cherniak, C., 1986. *Minimal Rationality*. Cambridge, MA: MIT Press.

Cohen, L.J., 1981. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317–370.

Cosmides, L. and Tooby, J., 1992. Cognitive adaptations for social exchange. *In:* J. Barkow, L. Cosmides, and J. Tooby, eds., *The Adapted Mind: Evolu-*

tionary Psychology and the Generation of Culture. Oxford: Oxford University Press, 163–228.

Cosmides, L. and Tooby, J., 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, 58, 1–73.

Dennett, D.C., 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.

Fiedler, K., 1988. The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123–129.

Foley, R., 1987. *The Theory of Epistemic Rationality*. Cambridge, MA: Harvard University Press.

Gigerenzer, G., 1991a. How to make cognitive illusions disappear: Beyond heuristics and biases. *European Review of Social Psychology*, 2, 83–115.

Gigerenzer, G., 1991b. On cognitive illusions and rationality. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 21, 225–249.

Gigerenzer, G., 1994. Why the distinction between single-event probabilities and frequencies is important for psychology (and vice-versa). *In:* C. Wright and P. Ayton, eds., *Subjective Probability*. New York: John Wiley, 129–161.

Gigerenzer, G., 1998. Ecological intelligence: An adaptation for frequencies. *In:* D.D. Cummins and C. Allen, eds., *The Evolution of Mind*. Oxford: Oxford University Press, 9–29.

Gigerenzer, G., 2000. *Adaptive Thinking: Rationality in the Real World*. Oxford: Oxford University Press.

Gigerenzer, G. and Goldstein, D.G., 1996. Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 102, 650–669.

Gigerenzer, G. and Hoffrage, U., 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684–704.

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

Goldberg, L., 1965. Diagnosticians vs. diagnostic signs: The diagnosis of psychosis vs. neurosis from the MMPI. *Psychological Monographs*, 79, 1–28.

Golding, J., Limerick, S. and Macfarlane, A., 1985. *Sudden Infant Death*. Seattle: University of Washington Press.

Goldman, A.I., 1978. Epistemics: The regulative theory of cognition. *Journal of Philosophy*, 75, 509–523.

Goldman, A.I., 1979. What is justified belief?. *In:* G. Pappas, ed., *Justification and Knowledge*. Dordrecht: Reidel, 1–23.

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

Goldman, A.I., 1993. *Philosophical Applications of Cognitive Science*. Boulder, CO: Westview.

Griggs, R. and Cox, J., 1982. The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73, 407–420.

Harman, G., 1973. *Thought*. Princeton, NJ: Princeton University Press.

Harman, G., 1995. Rationality. *In:* E.E. Smith and D.N. Osherson, eds., *Thinking*. Cambridge, MA: MIT Press, 175–211.

Hertwig, R. and Gigerenzer, G., 1994. The chain of reasoning in the conjunction task. *Unpublished manuscript*.

Kahneman, D. and Tversky, A., 1973. On the psychology of prediction. *Psychological Review*, 80, 237–251.

Kahneman, D., Slovic, P. and Tversky, A., 1982 (eds.). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

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

Nisbett, R. and Borgida, E., 1975. Attribution and the social psychology of prediction. *Journal of Personality and Social Psychology*, 32, 932–943.

Nozick, R., 1993. *The Nature of Rationality*. Princeton, NJ: Princeton University Press.

Osherson, D.N., 1995. Probability judgment. *In:* E.E. Smith and D.N. Osherson, eds., *Thinking*. Cambridge, MA: MIT Press, 33–75.

Samuels, R. and Stich, S., 2004. Rationality and psychology. *In:* A. Mele and P. Rawlings, eds., *The Oxford Handbook of Rationality*. Oxford: Oxford University Press, 279–300.

Samuels, R., Stich, S. and Bishop, M., 2002. Ending the rationality wars: How to make the disputes about human rationality disappear. *In:* R. Elio, ed. *Common Sense, Reasoning and Rationality*. Oxford: Oxford University Press, 236–268.

Simon, H.A., 1957. *Models of Man*. New York: Wiley.

Stanovich, K.E., 1999. *Who Is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum.

Stich, S.P., 1990. *The Fragmentation of Reason*. Cambridge, MA: MIT Press.

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.

Tversky, A. and Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.

Wason, P.C., 1966. Reasoning. *In:* B. Foss, ed., *New Horizons in Psychology*. Harmonsworth, England: Penguin, 135–151.