Epistemology and the Psychology of Human Judgment

Michael A Bishop

J. D. Trout

Introduction

The first three chapters of the book introduce the basic building blocks of our epistemological approach and of our epistemological theory. Chapter 1 introduces the basic motives and methods of our epistemology. The goal is to give the reader a clear conception of our overall project. As a result, the opening chapter is not weighed down with arguments and qualifications—that comes later. Chapter 2 introduces Statistical Prediction Rules (SPRs) and offers an explanation for their success. SPRs are simple, formal rules that have been shown to be at least as reliable, and typically more reliable, than the predictions of human experts on a wide variety of problems. On the basis of testable results, psychology can make *normative recommendations* about how we ought to reason. We dub the branches of psychology that provide normative recommendations "Ameliorative Psychology." Ameliorative Psychology recommends SPRs on the basis of testable results: SPRs are reliable, they tend to be easy to use, and they typically address significant issues. In addition, taking seriously the success of SPRs requires us to impose discipline on a human mind that is much too easily tempted by appealing distractions. Certain lines of evidence, no matter how subjectively attractive or how consecrated by the concepts central to our epistemological tradition, are to be ignored except in extreme cases. In chapter 3, we identify some of the basic building blocks of the epistemological framework that supports the recommendations of Ameliorative Psychology. The framework assesses the epistemic merit of reasoning strategies in terms of their robust reliability, their feasibility and the significance of the problems they tackle. We then argue

that this framework offers a new way to think about applied epistemology. In particular, it suggests that there are four and only four ways for people to improve their reasoning.

The middle three chapters of the book (4, 5, and 6) articulate the central features of our theory of epistemic excellence, Strategic Reliabilism. Strategic Reliabilism holds that epistemic excellence involves the efficient allocation of cognitive resources to robustly reliable reasoning strategies applied to significant problems. Chapter 4 takes up what is the central notion of Strategic Reliabilism: what it is for a reasoning strategy to be robustly reliable. Chapter 5 defends a cost-benefit approach to epistemology and offers an account of what it is for cognitive resources to be allocated efficiently. Chapter 6 argues that a genuinely normative epistemological theory must include some notion of significance, and it addresses the issue of what it is for a problem to be significant.

The final three substantive chapters of the book (7, 8, and 9) put our views about epistemology to work. In chapter 7, we criticize the approach that has dominated English-speaking philosophy over the past half-century or so—what we call Standard Analytic Epistemology (SAE). SAE names a contingently clustered set of methods and motives. By comparing our approach to that of SAE, chapter 7 identifies some of the troubles with SAE and argues that they are serious enough to motivate a radically different approach to epistemology. Chapter 8 takes Strategic Reliabilism, which has been extracted from psychology, and turns it back on psychology. We use Strategic Reliabilism to resolve two debates about whether certain experimental findings demonstrate deep and systematic failures of human reasoning. This chapter illustrates one of the main benefits of our approach to epistemology: it can be used to adjudicate disputes that arise in psychology that are, at bottom, normative epistemological disputes about the nature of good reasoning. Chapter 9 attempts to consolidate some of the lessons of Ameliorative Psychology with some handy heuristics and illustrative injunctions. We explore the empirical research that shows how we can enhance the accuracy of diagnostic reasoning, reduce overconfidence, avoid the regression fallacy, improve our policy assessments, and restrain the unbridled story-telling surrounding rare or unusual events. We have no doubt that many significant problems we face are best addressed with institutional measures, and on this issue much research remains to be done. But for those problems tractable to voluntary reasoning strategies, the simple strategies recommended in this chapter can improve reasoning at low cost and high fidelity.

Chapter 10 briefly sums up our view and points to some of the challenges that remain in the construction of a naturalistic epistemology.

The Appendix considers 11 objections that we expect philosophers to level against our views. Some will undoubtedly complain that we have missed some serious objections, or that our replies to the objections we do consider are by no means conclusive. Granted. But our goal in the Appendix is not the wildly ambitious one of overcoming all serious objections. Instead, our aim is to offer some sense of the resources available to the naturalist for overcoming what many proponents of SAE are likely to consider devastating objections.

Laying Our Cards
on the Table

It is time for epistemology to take its rightful place alongside ethics as a discipline that offers practical, real-world recommendations for living. In our society, the powerful are at least sometimes asked to provide a moral justification for their actions. And there is at least sometimes a heavy price to be paid when a person, particularly an elected official, is caught engaging in immoral actions or defending clearly immoral policies. But our society hands out few sanctions to those who promote and defend policies supported by appallingly weak reasoning. Too often, condemnation is meted out only after the policies have been implemented and have led to horrible results: irresponsible war and spilt blood or the needless ruin of people's prospects and opportunities.

Epistemology is a serious business for at least two reasons. First, epistemology guides reasoning, and we reason about everything. If one embraces a defective morality, one's ability to act ethically is compromised. But if one embraces a defective epistemology, one's ability to act effectively in all areas of life is compromised. Second, people don't fully appreciate the risks and dangers of poor reasoning. Everyone knows the danger of intentional evil; but few fully appreciate the real risks and untold damage wrought by apparently upstanding folk who embrace and act on bad epistemological principles. Such people don't look dangerous. But they are. An example of the costs of upstanding people reasoning poorly is the surprisingly strong opposition in the United States to policies that would provide opportunities and services to the disadvantaged (e.g., in terms of education and

basic needs such as health care). Much of this opposition is not based on the rejection of a moral principle of equal opportunity, but instead on poorly-arrived-at empirical views. Some people reject redistributive social policies on the grounds that they are *inevitably* ineffective; others rely on clearly mistaken views about what percentage of the federal budget actually goes to pay for such programs. That's not to say that there aren't good arguments against some redistributive policies. Some can backfire, and others (particularly those that benefit the non-poor) can be very expensive. But sound comparative policy analysis provides no support to a *principled* opposition to redistributive social policies. People who defend appalling social policies often do so on the basis of weak reasoning about factual matters rather than on the basis of backward moral precepts.

One might think that our call for a more prescriptive, reason-guiding epistemology is more appropriate for the areas of "critical thinking" or "informal logic" (Feldman 1999, 184–85, n10). The problem with this suggestion is that these areas, as exemplified in textbooks, are completely divorced from contemporary epistemology. This bespeaks deep problems both for critical thinking courses and for contemporary epistemology. Epistemology, if it is to achieve its normative potential, must make firm contact with the sorts of reasoning errors that lead to horrendous and avoidable outcomes. And critical thinking courses must be informed by a theory about what makes reasoning good or bad. We do not have in mind a thin epistemological "theory" (e.g., "premises should be true and support the conclusion") that yields a long list of informal fallacies. Rather, an effective critical thinking course should be informed by a theory that (among other things) helps us to recognize, anticipate, and compensate for our cognitive frailties. In other words, such courses should be informed by a deeply naturalistic epistemological theory.

We have written this book driven by a vision of what epistemology could be—normatively reason guiding and genuinely capable of benefiting the world. If our tone is not always dispassionate, it is because our profession has so clearly failed to bring the potential benefits of epistemology to ordinary people's lives. We are under no illusions, however. This book is, at best, a modest first step toward the construction of an epistemological theory with concrete, prescriptive bite. And even if our theory should be somewhere close to the truth, we are not sanguine about the potential of philosophy to influence the world. Sometimes, though, life rewards wild-eyed optimists. If in our case it doesn't, we fall squarely within what is best in our philosophical tradition if our reach should exceed our grasp.

1. Starting points: What epistemology is about

Theories, including epistemological theories, are supposed to be about something. They are supposed to explain or account for some range of phenomena. An important way in which our approach to epistemology differs from that of most contemporary English-speaking epistemologists is in terms of what we take to be the proper subject matter of epistemology—what we take to be the phenomena or evidence that an epistemological theory is supposed to account for or explain. Traditional epistemological theories aim to provide a theory that captures our considered epistemic judgments, in particular, our considered judgments about knowledge and justification. Our epistemological theory aims to uncover the normative assumptions of a branch of science. We disagree with most traditional epistemologists in terms of what epistemology is about. This difference couldn't be more fundamental.

1.1. The starting point of the standard analytic approach to epistemology

Standard Analytic Epistemology (SAE) names a contingently clustered class of methods and theses that have dominated English-speaking epistemology for much of the past century. Almost all the contemporary readings in the most popular epistemology textbooks are prime examples of SAE. Contemporary versions of foundationalism, coherentism, and reliabilism are exemplars of SAE. While we object to the methods of SAE, and therefore to the kinds of theories it leads to, our main goal in this chapter is to distinguish our approach from that of SAE. So let's begin with the starting points of SAE—what proponents of SAE take to be the fundamental phenomena or evidence of epistemology.

The goal of most philosophers engaged in SAE is to provide an account of knowledge and epistemic justification. What are the *success conditions* on such an account? In a typically clear and careful article, Jaegwon Kim identifies a number of criteria that any account of justification must meet in order to be successful. The most important of these conditions is what we will call the *stasis requirement*:

> Although some philosophers have been willing to swallow skepticism just because what we regard as correct criteria of justified belief are seen to lead inexorably to the conclusion that none, or very few, of our beliefs are

justified, the usual presumption is that our answer to the first question [What conditions must a belief meet if we are justified in accepting it as true?] should leave our epistemic situation largely unchanged. That is to say, it is expected to turn out that according to the criteria of justified belief we come to accept, we know, or are justified in believing, pretty much what we reflectively think we know or are entitled to believe. (Kim 1988, 382)

It is worth noting that this requirement—that the right account of justification "leave our epistemic situation largely unchanged"—is profoundly conservative. In particular, it is extraordinary that SAE should have *built right into it* a requirement that makes it virtually impossible that a successful epistemological theory would force us to radically alter our epistemic judgments.

Of course, proponents of SAE will not suggest that they are trying to provide an account of their naïve epistemic judgments, but of their considered epistemic judgments. One way to spell out the difference is in terms of reflective equilibrium. Nelson Goodman introduced reflective equilibrium as a process that involves aligning our judgments about particular instances with our judgments about general principles. "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). Narrow reflective equilibrium is the process of bringing our normative judgments about particular cases into line with our general normative prescriptions and vice versa. Wide reflective equilibrium differs from narrow reflective equilibrium by including our best theories in the mix. So wide reflective equilibrium is the process of bringing into alignment our best theories, as well as our normative judgments about particular cases, and our general normative prescriptions (Rawls 1971, Daniels 1979).

So according to the stasis requirement, if an epistemic theory forced us to radically alter our considered epistemic judgments (e.g., our epistemic judgments in reflective equilibrium), then ipso facto that theory is unacceptable. While some proponents of SAE might reject the stasis requirement (e.g., Unger 1984), we agree with Kim that stasis is a fundamental commitment of SAE. It is not, however, often explicitly stated. That is because the commitment to epistemic stasis is implicit in the practice of SAE. Much of SAE proceeds by counterexample philosophy: Someone proposes an account of justification, others propose counterexamples, and then the original account is altered or defended in the face of those counterexamples. What we find objectionable about this mode of argument is what proponents of SAE accept as a *successful* counterexample. To see

this, let's consider the mother-of-all counterexamples in SAE, the Gettier Problem.

Before Gettier, it was generally thought that knowledge is justified true belief (JTB). Gettier (1963) describes a situation in which the JTB account is at odds with our considered knowledge judgments. One of Gettier's famous cases involves a man named Smith who has overwhelming evidence, and so justification, for believing that Jones will get a job and that Jones has ten coins in his pocket. On the basis of these beliefs, Smith infers that the man who will get the job has ten coins in his pocket. It turns out that unbeknownst to Smith, he will get the job, and he has ten coins in his pocket. His belief that the man who will get the job has ten coins in his pocket is true and justified. But Gettier insists that it is "clear" that Smith's belief is not knowledge (Gettier 1963, 122). For proponents of SAE, the Gettier examples are important because they show that the JTB account can't be right on the grounds that it does not "leave our epistemic situation largely unchanged." Rather than explore any more of the countless and wonderfully rococo counterexamples prevalent in the SAE literature, let's look at how some of these counterexamples end:

> However, it is perfectly apparent that I know nothing of the sort. (Lehrer and Paxson 1969, 235)

> Even if S correctly predicts that he is going to lose, we would deny that he knew he was going to lose if the only basis he had for this belief was the fact that his chances of winning were so slight. (Dretske 1971, 3)

> The situation is a peculiar one, and my intuitions, and I would suppose other people's, are not completely clear on the matter. But it seems, on the whole, that we ought not to speak of knowledge here. . . . (Armstrong 1973, 181)

> But, to make such an assumption is counterintuitive. In everyday situations we do not regard deception as precluding rationality. Likewise, we do not regard the fact that we have been deceived, or will be deceived, or would be deceived, as precluding rationality. (Foley 1985, 192)

> And, surely, we do not want to say that the fact that his friend has a generator in his basement *prevents* S from having knowledge that the company's generators are causing the lights to be on. (Pappas and Swain 1973, 66)

In the above passages (and we could have chosen literally hundreds of others), we are urged to share the philosopher's considered epistemic judgments about some imagined scenario. And we usually do. The problem,

on our view, is that SAE rejects various accounts *solely on the grounds that they violate these judgments.*

The shockingly conservative nature of the method of SAE may only become clear when we compare it to methods in other fields of inquiry. The fact that relativity denies people's considered judgments about simultaneity is hardly a reason to reject it. If physics had been burdened with such a conservative method, we wouldn't have relativity, quantum mechanics (or perhaps even Copernicanism!). If biology had been taken over by such a conservative method, we wouldn't have Darwinism. If cultural studies had had such a conservative method, we wouldn't have postmodernism.

Okay, so sometimes conservatism is a good thing.

Behind this joke is an important point. The problem with conservative methods is not that they are conservative per se. Conservative methods work very well when applied to theories or propositions for which we have overwhelming evidence. It is perfectly reasonable to be conservative about the commitments of theoretical chemistry reflected in the periodic table, or about the core attachments of contemporary physics or biology. That doesn't mean we rule out the possibility that new developments will force us to abandon them. Conservatism isn't mulishness. Conservatism is appropriate in the case of the core commitments of these theories because we have so much evidence in their favor that in absence of extraordinary counterevidence, they deserve our allegiance. But while conservatism is fine for excellent theories, it is poison in domains where progress awaits deep and durable changes in method and outlook. The alchemist's attachment to conservatism was ill advised; it only protracted the alchemist's crippling (and it turns out, thanks to mercury and lead, fatal) ignorance. This raises an obvious concern for SAE, which we will explore more fully in chapter 7. No matter how polished or well thought-out our epistemic judgments, no matter how much in reflective equilibrium they might be, are we so confident in them that it is reasonable to make them the final arbiters of our epistemological theories?

1.2. The starting point of the philosophy of science approach to epistemology

We view epistemology as a branch of the philosophy of science. From our perspective, epistemology begins with a branch of cognitive science that investigates good reasoning. It includes work in psychology, statistics,

machine learning, and Artificial Intelligence. Some of this work involves "predictive modeling," and it includes discussion of models such as linear models, multiple regression formulas, neural networks, naïve Bayes classifiers, Markov Chain Monte Carlo algorithms, decision tree models, and support vector machines; but much of this work comes from traditional psychology and includes the well-known heuristics and biases program launched by Kahneman and Tversky (Kahneman, Slovic, and Tversky 1982). It will be useful to give this wide-ranging literature a name. We call it *Ameliorative Psychology*. The essential feature of Ameliorative Psychology is that it aims to give positive advice about how we can reason better. We will introduce many findings of Ameliorative Psychology (particularly in chapters 2 and 9). But it will be useful here to introduce some of its noteworthy features.

In the course of this book, we will introduce a number of reason-guiding prescriptions offered by Ameliorative Psychology. This advice includes making statistical judgments in terms of frequencies rather than probabilities, considering explanations for propositions one doesn't believe, ignoring certain kinds of evidence (e.g., certain selected cues that improve accuracy only very moderately, and certain kinds of impressionistic information, such as opinions gleaned from unstructured personal interviews), and many others (Bishop 2000). These recommendations are bluntly normative: They tell us how we *ought* to reason about certain sorts of problems.

A particularly interesting branch of Ameliorative Psychology begins in earnest in 1954 with the publication of Paul Meehl's classic book *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Meehl reported on twenty experiments that showed that very simple prediction rules were more reliable predictors than human experts. Since then, psychologists have developed many of these Statistical Prediction Rules (or SPRs). (In fact, in the past decade or so, there has been an explosion of predictive models in AI and machine learning.) There is now considerable evidence for what we call *The Golden Rule of Predictive Modeling*: When based on the same evidence, the predictions of SPRs are at least as reliable, and are typically more reliable, than the predictions of human experts. Except for an important qualification we will discuss in chapter 2, section 4.2, the evidence in favor of the Golden Rule is overwhelming (see Grove and Meehl 1996; Swets, Dawes, and Monahan 2000).

The Golden Rule of Predictive Modeling has been woefully neglected. Perhaps a good way to begin to undo this state of affairs is to briefly

describe ten of its instances. This will give the reader some idea of the range and robustness of the Golden Rule.

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

2. A model that used past criminal and prison records was more reliable than expert criminologists in predicting criminal recidivism (Carroll et al., 1988).

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

4. A number of SPRs predict academic performance (measured by graduation rates and GPA at graduation) better than admissions officers. This is true even when the admissions officers are allowed to use considerably more evidence than the models (DeVaul et al. 1957), and it has been shown to be true at selective colleges, medical schools (DeVaul et al. 1957), law schools (Swets, Dawes, and Monahan 2000, 18), and graduate school in psychology (Dawes 1971).

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

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

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

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

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

10. In legal settings, forensic psychologists often make predictions of violence. One will be more reliable than forensic psychologists simply by

predicting that people will not be violent. Further, SPRs are more re-
liable than forensic psychologists in predicting the relative likelihood of
violence, that is, who is more prone to violence (Faust and Ziskin 1988).

Upon reviewing this evidence in 1986, Paul Meehl said: "There is no
controversy in social science which shows such a large body of qualita-
tively diverse studies coming out so uniformly in the same direction as this
one. When you are pushing [scores of] investigations, predicting every-
thing from the outcomes of football games to the diagnosis of liver disease
and when you can hardly come up with a half dozen studies showing even
a weak tendency in favor of the clinician, it is time to draw a practical
conclusion" (Meehl 1986, 372–73). Ameliorative Psychology has had
consistent success in recommending reasoning strategies in a wide variety
of important reasoning tasks. Such success is worth exploring.

The descriptive core of our approach to epistemology consists of the
empirical findings of Ameliorative Psychology. And yet, Ameliorative
Psychology is deeply normative in the sense that it makes (implicitly or
explicitly) evaluative "ought" claims that are intended to guide people's
reasoning. Let's look at three examples of the reason-guiding prescriptions
of Ameliorative Psychology.

A well-documented success of Ameliorative Psychology is the Gold-
berg Rule (the third item on the above list). It predicts whether a psy-
chiatric patient is neurotic or psychotic on the basis of an MMPI profile.
Lewis Goldberg (1965) found that the following rule outperformed 29
clinical judges (where L is a validity scale and Pa, Sc, Hy, and Pt are clinical
scales of the MMPI):

$$x = (L + Pa + Sc) - (Hy + Pt)$$

If $x < 45$, diagnose patient as neurotic.

If $x \geq 45$, diagnose patient as psychotic.

When tested on a set of 861 patients, the Goldberg Rule had a 70% hit rate;
clinicians' hit rates varied from a low of 55% to a high of 67%. (13 of the 29
clinical judges in the above study were experienced Ph.D.s, while the other
16 were Ph.D. students. The Ph.D.s were no more accurate than the stu-
dents. This is consistent with the findings reported in Dawes 1994.) So here
we have a prediction rule that could literally turn a smart second-grader
into a better psychiatric diagnostician than highly credentialed, highly
experienced psychologists—at least for this diagnostic task. In fact, more
than 3 decades after the appearance of Goldberg's results, making an initial

diagnosis on the basis of an MMPI profile by using subjective judgment rather than the Goldberg Rule would bespeak either willful irresponsibility or deep ignorance. So here is a finding of Ameliorative Psychology: people (in an epistemic sense) *ought* to use the Goldberg Rule in making preliminary diagnoses of psychiatric patients.

Another example of Ameliorative Psychology making evaluative ought-claims is a 1995 paper by Gigerenzer and Hoffrage entitled "How to *Improve* Bayesian Reasoning Without Instruction: Frequency Formats" (emphasis added). As the title of the paper suggests, Gigerenzer and Hoffrage show how people charged with making high-stakes diagnoses (e.g., about cancer or HIV) can *improve* their reasoning. They suggest a reasoning strategy that enhances reasoners' ability to identify, on the basis of medical tests, the likelihood that an individual will have cancer or HIV. We will discuss these "frequency formats" in chapter 9, section 1. For now, it is enough to note that a finding of Ameliorative Psychology is that people *ought* to use frequency formats when diagnosing rare conditions on the basis of well-understood diagnostic tests.

Another particularly successful example of Ameliorative Psychology is credit scoring (the fifth item on the above list). Many financial institutions no longer rely primarily on financial officers to make credit decisions—they now make credit decisions on the basis of simple SPRs developed as the result of research by psychologists and statisticians (Lovie and Lovie 1986). Once again, this finding of Ameliorative Psychology seems to be normative through and through: When it comes to making predictions about someone's creditworthiness, one *ought* to use a credit-scoring model.

Not only does Ameliorative Psychology recommend particular reasoning strategies for tackling certain kinds of problems, it also suggests generalizations about how people ought to reason. (See, for example, the flat maximum principle, discussed in chapter 2, section 2.1.) On our view, the goal of epistemology is to articulate the epistemic generalizations that guide the prescriptions of Ameliorative Psychology. In this way, epistemology is simply a branch of philosophy of science. Just as the philosopher of biology might aim to uncover and articulate the metaphysical assumptions of evolutionary theory, the epistemologist aims to uncover and articulate the normative, epistemic principles behind the long and distinguished tradition of Ameliorative Psychology. (There are two objections philosophers are likely to immediately raise against our approach. We consider them in the Appendix, sections 1 and 2.)

Ameliorative Psychology is normative in the sense that it yields explicit, reason-guiding advice about how people ought to reason. Some

might fix us with a jaundiced eye and wonder whether the recommendations of Ameliorative Psychology are really normative in the same way as the recommendations of SAE are normative. Admittedly, there does seem to be one telling difference. People outside academia have on occasion actually changed the way they reason about significant matters as a result of the recommendations of Ameliorative Psychology.

2. The end points: The theories generated by the two approaches

The two approaches to epistemology we have been considering differ in terms of what they take to be the appropriate subject matter of epistemology (our considered judgments vs. Ameliorative Psychology). Given that they differ so fundamentally regarding what epistemology is about, it is not surprising that they end up with quite different normative theories. Indeed, they end up with theories of different phenomena. The fundamental aim of SAE is to deliver an account of epistemic justification or knowledge (or one of their close relatives, e.g., warrant). The fundamental aim of our approach to epistemology is to provide an account of *reasoning excellence.* Is this really a deep difference? Yes, it is.

Justification, the target of theories of SAE, is a property of belief tokens. Judy might be justified in believing that George is a dolt, while Mary is not. So a theory of SAE will provide an account that distinguishes the justified belief tokens from the unjustified belief tokens (or, perhaps, the more justified belief tokens from the less justified belief tokens). Epistemic excellence, the target of our theory, we take to be a property of reasoning strategies. The primary normative assessments made by Ameliorative Psychology are of ways of reasoning. Ameliorative Psychology is in the business of telling us what are the best ways to go about (say) making tentative diagnoses of psychiatric patients (Goldberg Rule) or making judgments about a person's ability to repay a loan (credit-scoring models). So an epistemology that puts Ameliorative Psychology at center stage will yield a theory of reasoning excellence (see also Goldman 1979, Stich 1990).

While the notion of epistemic excellence might not have the common currency or the philosophical pedigree of notions like justification, rationality, or reason, it is a very useful concept to have at the center of one's epistemology. When a thoughtful person is faced with a reasoning problem, she will sometimes think about and try to figure out what is the best way to tackle this problem. We often have a sense (though perhaps sometimes

a mistaken sense) that certain reasoning strategies are better than others for handling certain reasoning problems. For example, deciding whether a prisoner up for parole is a threat to society on the basis of his record in prison is better than flipping a coin (and, as it turns out, not as good as using a decision tree; see Quinsey, et al. 1998). So we understand that some reasoning strategies are better than others; and often there is a reasoning strategy that is the best available. Our epistemological theory aims to provide an account of what it is for a reasoning strategy to be excellent, or better than any of the alternatives.

If our theory and the theories of SAE are theories of different episte-mological categories, one might wonder whether they can conflict. Perhaps by so radically altering what we take epistemology to be, we have changed the subject? We don't think this is a serious worry. A theory of justification will yield normative conclusions about *belief tokens*—whether they are justified or not (or the degree to which they are justified). A theory of epistemic excel-lence will yield normative conclusions about the epistemic quality of a *rea-soning strategy*. But reasoning strategies typically produce belief tokens. So whenever a theory of reasoning excellence recommends a particular rea-soning strategy for tackling a particular problem, it normally recommends a belief token, but at one remove. And this leaves open the possibility of conflict. It is possible for a theory of reasoning excellence to recommend a reasoning strategy to S that yields the belief that p, and for a theory of justification to conclude that S's belief that not-p is justified and that S's belief that p is not justified. Insofar as the two approaches to epistemology are meant to guide reasoning, it is possible for them to yield recommendations that are mutually incompatible (in the sense that both cannot be followed).

3. The structure of a healthy epistemological tradition

On our approach to epistemology, a healthy epistemological tradition must have three vigorous and interrelated components: theoretical, prac-tical, and social. The practical or applied component of epistemology is an extension of what people do every day. Everyone who has ever thought about how to tackle a particular reasoning problem has engaged in applied epistemology. As is standard with an applied venture, some people do it better than others. Ameliorative Psychology is the science of applied epis-temology. Much of the point of Ameliorative Psychology is to provide advice that will help people reason better about the world.

The second component of a healthy epistemological tradition is theory. We take theory and application to be mutually informing and supporting. Theory is extracted from practice. One of the goals—and one of the tests—of a theory of reasoning excellence is that it should be faithful to the practice of Ameliorative Psychology. When conjoined with the descriptive results of Ameliorative Psychology, the correct epistemological theory should yield the recommendations of Ameliorative Psychology. One of our primary goals in this book is to offer a theory that accurately depicts the normative machinery that guides the prescriptions of Ameliorative Psychology. But theory should do more than mimic. It should *explain* what makes some reasoning strategies epistemically better than others; it should also play a role in a full explanation for why good reasoning tends to lead to good outcomes. (To see how our theory addresses these explanatory challenges, see Appendix, section 8.) Further, a theory of reasoning excellence should be able to be applied back to Ameliorative Psychology. Practice informs theory; but good theory repays the kindness. When a disagreement erupts in the applied domain, and that disagreement is at bottom a theoretical one, a good theory should be able to clarify and, in some cases at least, resolve the issues. In chapter 8, we will apply our theory of reasoning excellence in an effort to resolve two disputes that have arisen in Ameliorative Psychology.

We have suggested that the theoretical part of a healthy epistemological tradition will be firmly connected to its applied components. As we have already suggested, by this yardstick, the standard analytic approach to epistemology does not seem to be a healthy tradition. As far as we have been able to tell, the theoretical musings of analytic epistemologists have not led to very much, if any, useful guidance about how people should reason. We will argue eventually that this prescriptive impotence is a natural consequence of the methods of Standard Analytic Epistemology. If this is right, it is a shame. It is the normative, reason-guiding promise of epistemology that makes it so much more than intellectual sport.

While a healthy epistemological tradition will provide useful reasoning guidance, good advice we keep to ourselves is no advice at all. Ameliorative Psychology is the science of applied epistemology, and theoretical epistemology is theoretical Ameliorative Psychology (i.e., a theoretical science). As with any science, it is important to think about what it would take for it to be a well-ordered social system (Kitcher 2001). An important aspect of epistemology's social presence is how it communicates its practical recommendations to the wider public. We don't have any detailed picture of what a socially well-ordered epistemology would look like. But we are

confident that it would have at least two features. First, in order to achieve its ameliorative potential, epistemology should be organized so that it provides a way to effectively communicate its established findings, particularly its practical content, to a wide audience. Second, in order to minimize the risk of promulgating harmful or mistaken findings, epistemology should be organized so that whatever findings are communicated widely will have passed rigorous examination and empirical testing.

4. Seductive circularities and empirical hooks: Is a scientific investigation into normative epistemology possible?

We have argued that applied epistemology is a science, and that theoretical epistemology is a theoretical science. But we also seek an epistemic theory that is normative and reason guiding. How can a scientific epistemology also be a normative one? The standard worry with our approach is that it is somehow viciously circular. The objection goes like this: Suppose our epistemological theory begins with empirical claims about Ameliorative Psychology. Presumably, we have to make some decisions about which empirical claims to trust. So we have to decide which views are the epistemically good ones. But such decisions require a prior epistemological theory. So (the argument continues) one cannot begin one's epistemological speculations with empirical claims. (For a discussion of this objection, see Appendix, section 2.)

   This is a very seductive argument. One problem with it is that it assumes the normative must come in a single dollop. So either one has a full-blown theory and can make normative judgments or one has no theory and can make no such judgments. If knowledge of the normative were an all-or-nothing affair, then a scientific epistemology, one that began with, say, Ameliorative Psychology, might be impossible. But it's not. In fact, Aristotle points the way to avoiding the theoretical stultification that comes with the dollop assumption.

   Aristotle argued that at least some of the moral and the intellectual virtues are intimately related and mutually supportive (*Nicomachean Ethics*, Book VI). Aristotle's insight provides us with an empirical ''hook'' into our investigation of the normative. To see how this hook works, suppose we're faced with making parole decisions for people convicted of a violent crime. An important question to consider is whether the prisoner is likely to commit another violent crime. Suppose we decide to use the Shoe

Size Rule: If the prisoner's shoe size is a whole number (e.g., 9, 10, 11), he won't commit another violent crime; if it's not (e.g., 9½, 10½), he will commit another violent crime. The Shoe Size Rule is a poor reasoning strategy. And there is a tell-tale empirical mark of its being a poor reasoning strategy: In the long run, the Shoe Size Rule will lead to poor outcomes—or more precisely, it will lead to worse outcomes than better reasoning strategies. Now, this notion of bad outcomes is not particularly subtle or in need of philosophical elucidation. Reasoning poorly about this problem will lead to increases in murder and assault by paroled prisoners. Similarly, if medical doctors reason poorly about whether patients have brain damage, cancer, or HIV, patients will tend to have worse treatment outcomes. Again, this isn't a particularly subtle point. Poor reasoning in these matters will lead patients to make treatment decisions that will lead to unnecessary death, suffering, and illness. (More precisely, poor reasoning will tend to lead to worse outcomes for patients than will good reasoning.)

The Aristotelian Principle says simply that *in the long run, poor reasoning tends to lead to worse outcomes than good reasoning*. So the Aristotelian Principle allows us to empirically determine—though not with complete certainty—when one way of reasoning is better than another. Of course, there are no guarantees. It is logically possible for someone to have bad luck and for terrific reasoning to lead consistently to bad outcomes; and it is logically possible for someone to reason badly and yet, Magoo-like, to have consistently good outcomes. But seldom does anything good in life come with guarantees. To begin our empirical investigation into the epistemological, all we really need is the robust generalization we have called the Aristotelian Principle. It allows us to accept certain normative epistemological judgments as prima facie true and then explore more deeply the sorts of assumptions that drive such judgments. This is how we will start our investigations into the normative.

Why should anyone believe the Aristotelian Principle? It is an empirical, probabilistic claim and, as such, it is child's play to imagine environments that are so unfriendly as to make excellent reasoning a danger (e.g., a powerful evil demon sets out to punish excellent reasoners). But as a practical matter, we contend that any psychologically healthy, reflective person who has chosen to spend their life doing epistemology must accept the Aristotelian Principle. It is a necessary precondition for the practical relevance of epistemology. Recall that we opened this chapter by arguing that epistemology is important because it has real potential to improve people's lives. The Aristotelian Principle embodies this promise. If the Aristotelian Principle is false, if good reasoning doesn't tend to lead to better

outcomes than bad reasoning, then epistemology can't be practically important. It would be like the New York Times crossword puzzle: an intellectual challenge, perhaps even an addictive one, but nothing more than an amusing pastime. More importantly, however, if the Aristotelian Principle is false, then we can't know how to lead our cognitive lives. Suppose we have to reason badly in order to achieve good outcomes. There are indefinitely many different ways to reason badly. And all of these ways of reasoning badly will typically lead to many, many different judgments about the world. Which way of reasoning badly will lead to good outcomes? Presumably we need to figure this out. But how are we supposed to figure that out? By reasoning well? Presumably not. But if by reasoning poorly, then once again, which way of reasoning poorly? And how are we supposed to figure that out? And so goes the infinite regress. . . .

If a useful epistemology is possible, then the Aristotelian Principle is true. But this raises an obvious and cynical worry: Is a useful epistemology really possible? There are at least two reasons for optimism. The first is that much of the world that is significant to us is stable enough for the quality of our reasoning to make a difference. We reason about medical diagnoses, policy choice, financial planning, criminal recidivism, etc. These (and many other) parts of the world have proven to be predictable enough for people to make judgments about them and make effective plans based on those judgments. The second reason to be optimistic about the Aristotelian Principle is that the human predicament comes with some stern and demanding contours. As people, we share substantial priorities. A good life, in general, will favor such things as health, shelter, satisfying, loving relationships, and the development of talents, interests, and other capabilities. Of course, there are myriad and surprising ways in which those facts can be realized. Our Aristotelian Principle does not depend on the Aristotelian view that the human ideal looks suspiciously like an ancient Greek philosopher (or a contemporary American one). A stable environment and the firm but multiply realizable boundaries of human welfare give us reason to be optimistic about the Aristotelian Principle and about the possibility of an effective, useful epistemology. Our goal in this book is to test this prospect.

5. Our uneasy relationship to tradition

When we began to study epistemology in graduate school, it seemed so full of promise. Who wouldn't want to divine the structure of knowledge? But

somewhere around the third epicycle on a counterexample involving re-
liable clairvoyants, back-up electrical generators, or an environment full of
objects that are phenomenologically identical but ontologically distinct,
SAE jumped the shark. ("Jumping the shark" is a specific allusion to the
episode—indeed, the moment—when Fonzie jumped the shark on the
sitcom Happy Days, in a shameless effort to resuscitate the failing sitcom.
It is a generic reference to any such moment in any TV series when it
becomes clear that the show is done for. People can disagree about when
or even whether a TV series has jumped the shark. The same goes for
advocates of particular philosophical movements.) At some point, we (and
we suspect at least some of our contemporaries) came to an uneasy and
perhaps not fully articulated realization that SAE is not what we signed up
for. It has taken us some time to put our finger on what we think the real
problem is. We think that the main problem with SAE is methodological:
its goals and methods are beyond repair. They guarantee that SAE will
never provide effective normative guidance, and so it will never achieve
the positive, practical potential of epistemology. In fact, we sometimes
despair about whether most contemporary epistemologists have lost sight
of this potential—and, indeed, of our obligation to seek it. We should
admit, however, that reliabilism has achieved some of epistemology's
reason-guiding potential. But as long as reliabilism remains wedded to the
goals and methods of SAE, it is doomed. That's because the real virtue of
reliabilism is not that it provides a perspicuous account of our concept of
justification. The real virtue of reliabilism lies in its reason-guiding (and
therefore action-guiding) potential.

Our perspective is uncompromisingly naturalistic. The standard phil-
osophical literature is full of questions and concerns about naturalism:
What is the appropriate way to formulate it? Does it entail that all
knowledge is third person? Does naturalism undermine first-person au-
thority? Is a fully naturalistic epistemology compatible with internalism, or
with externalism? Does it rule out epistemology's normative function? Is
naturalistic epistemology even possible? Inevitably, these issues get inte-
grated with metaphysical ones: Does naturalism entail materialism? Does
it entail reductionism? In the face of these worries, we can do no better
than to quote Elliott Sober: "Mark Twain once said that the trouble with
the weather is that everyone talks about it, but no one does anything about
it. I have had a similar gripe, from time to time, about the current vogue
for naturalism in philosophy" (1997, 549). In putting forth our positive
views, we intend to ignore concerns raised about naturalism except when
it suits our theoretical or narrative purposes. Questions about the nature

of naturalism are at this point premature. The right approach is to first build a naturalistic theory (or lots of them) and then noodle over what epistemological naturalism is like and what it entails.

There are a number of arguments from SAE that purport to show that naturalism in epistemology is impossible or self-refuting or self-undermining. We propose to ignore these arguments in putting forth our theory (although we do consider some of them in the Appendix). Some philosophers might wonder, with perhaps more than a hint of outrage, how we can justify blithely ignoring serious worries about our approach. Our decision to ignore such worries is a strategic one. Consider two points. First, arguments for rejecting a naturalistic approach to epistemology provide a positive reason for avoiding naturalism *only if* there is an alternative approach to epistemology that is more promising. But we contend that SAE embodies an approach that cannot fulfill the legitimate and essential practical ambitions of epistemology. In fact, given the failure of nonnaturalistic theories to offer anything in the way of useful reason guidance, it is high time to try something different. Our second point is that the history of science suggests that it is a mistake to wait for all objections to be met before proposing and defending a new, minority or unpopular theory. Naturalistic epistemology really *is* doomed if naturalists insist on attempting to defeat the Hydra-headed arguments for why it is doomed. When you're outnumbered and you want to show your theory is possible, proposing an actual theory is the best and probably only way to do it.