J. Expt. Theor. Artif. Intell. 15(2003)203–216

Taylor & Francis
Taylor & Francis Group

Epistemology's search for significance

MICHAEL A. BISHOP[1] and J. D. TROUT[2]

[1] *Department of Philosophy and Religious Studies, 402 Catt Hall, Iowa State University, Ames, IA 50011, USA*
e-mail: mikebish@iastate.edu
[2] *Department of Philosophy and the formerly Parmly Institute, Loyola University 6525 North Sheridan Road, Chicago, IL 60626, USA*
e-mail: jtrout@orion.it.luc.edu

Abstract. Epistemology is supposed to guide our reason. This means that epistemology is practically important because our beliefs so often play a decisive role in how we decide to act. We argue that in order to be effectively action-guiding, traditional epistemological theories must be supplemented with an account of significance—of what makes one reasoning problem significant and another reasoning problem insignificant. We sketch the beginning of an account of significance that employs Signal Detection Theory. We then argue that any internalist theory of justification is not going to fit smoothly with a plausible account of significance. This calls into question the power of internalist theories of justification to effectively guide reason and action.

Received July 2002; accepted December 2002

Contemporary analytic epistemology is obsessed with justification. To many, this obsession might well seem perfectly natural and appropriate. After all, epistemology is the study of knowledge, and knowledge and justification are intimately related. Justification (or warrant) is usually taken to be that property that knowledge has but that (mere) true belief lacks. But there are many other epistemological issues and concepts that are worth exploring (and that are occasionally investigated), including rationality, epistemic virtue, epistemic excellence and epistemic significance. So, this raises the question: why the obsession with justification? Perhaps the answer is no deeper than professional intertia—but we tend to favour a different answer. The reason epistemologists lavish so much attention on justification is that it plays a particularly central role in how we ought to guide our cognitive endeavours. Justification is the primary epistemic guide to judgement. Other things being equal, one ought to hold justified beliefs rather than unjustified beliefs.

This sort of response strikes us as exactly right. Epistemology is a serious business—or at least, it can and should be. But if epistemology really is a serious business, then guiding reason is less than half the battle. Guiding our reason and judgement is important primarily because our beliefs so often play a decisive role in how we decide to intervene in the world. If epistemology is to be practically efficacious, then it must be capable of doing more than merely guiding our reason.

Journal of Experimental & Theoretical Artificial Intelligence ISSN 0952–813X print/ISSN 1362–3079 online

DOI: 10.1080/0952813021000055153

It must be prepared to guide our action as well. (We are assuming that the adoption of a belief is not an action. For those who would deny this (and *we* might deny this), they should read 'action' in the rest of this paper as 'action other than the adoption of a propositional attitude'.) We think that it is an unwritten assumption of contemporary epistemology that theories of justification can be easily supplemented with machinery that will help guide action (not just reason). Our aim in this paper is to argue that there is reason to doubt this assumption. Much contemporary epistemology unnaturally sunders belief from action. As a result, many normative theories of reasoning will not be easily connected to normative theories of action.

1. The focus of standard analytic epistemology

Contemporary epistemology's relentless focus on justification makes good sense if it is the central reason-guiding epistemic notion. The reason for the infatuation with justification is that epistemology is supposed to be deeply relevant to what people ought to believe and how people ought to reason. There are many examples of epistemologists focusing on the reason-guiding nature of epistemology, knowledge and justification. For example, Pollock and Cruz take 'the fundamental problem of epistemology to be that of deciding what to believe' (1999: 14). BonJour explicitly takes justification to be a reason-guiding notion:

> [O]ne's cognitive endeavors are epistemically justified only if and to the extent that they are aimed at [truth], which means very roughly that one accepts all and only those beliefs which one has good reason to think are true. To accept a belief in the absence of such a reason, however appealing or even mandatory such acceptance might be from some other standpoint, is to neglect the pursuit of truth; such acceptance is, one might say, *epistemically irresponsible*. (1985: 8).

Three views of knowledge and justification have dominated epistemology over the past decades: foundationalism, coherentism and reliabilism. There are many variations on these views, but, for our purposes, a brief thumbnail sketch of 'vanilla' versions of these views will suffice. We want to highlight two features of these views that will emerge in later sections. First, the three traditional theories of justification are not particularly concerned with the costs and benefits of reasoning. As far as these theories are concerned, whether we spend our finite cognitive resources thinking about the length of our toenails or about life-or-death matters is immaterial. Whatever we choose to reason about, the traditional theories can tell us whether we have arrived at justified or unjustified beliefs—and that is it. That these theories ignore issues of significance will be a major theme of our paper. For now, we will note that the failure to take issues of significance into account is a bit odd if the obsession with justification is defended on the grounds that justification is reason-guiding—and certainly odd if it is supposed to be action-guiding. A plausibly reason-guiding theory must, after all, recommend that one (at least sometimes) think about significant matters. The second feature of these theories we want to highlight is really a feature of just two of the three theories. Foundationalism and coherentism are internalist theories of justification. Epistemic internalists hold that all the features of a belief that make it justified are cognitively accessible to the believer. A useful way to formulate internalism is as a view about *justifiers*—the facts that determine the justificational status of a belief (Alston 1989). For internalists, all justifiers must be cognitively accessible. We will argue that the attempt to make internalist theories of justification sensitive to issues of significance is going to be somewhat problematic.

Foundationalists hold that many beliefs are justified in terms of their relations to other beliefs. However, there is a special class of beliefs that are justified, but not in virtue of their relation to other beliefs. These basic beliefs are immediately or directly justified. The basic beliefs are typically taken to be perceptual beliefs. The non-basic beliefs are justified ultimately in terms of whether they are appropriately related to the basic beliefs. Traditionally, foundationalists have held that the basic beliefs include those given directly by experience. Critics of foundationalism have focused most of their attention on how beliefs can be immediately (or directly) justified. Notice that foundationalism says nothing about which justified beliefs one should adopt. As far as foundationalism is concerned, justified beliefs about toenail length are on all fours with justified beliefs about whether one has cancer.

Like foundationalists, *coherentists* hold that beliefs can be justified in terms of their relations to other beliefs, but coherentists deny the existence of basic beliefs. For coherentists, a belief is justified to the extent that it is part of a coherent belief-system. Perhaps the best way to understand coherentism is in terms of its response to the primary argument for foundationalism. The main premise of the regress argument is the following conditional: if knowledge does not have basic foundational beliefs, then all beliefs would require other beliefs to justify them, leading to an infinite regress. The foundationlist's tollens is the coherentist's ponens. In other words, from the conditional, the foundationalist infers that knowledge does have foundations, while the coherentist infers that knowledge does not have foundations and that all beliefs are justified in terms of their relations to other beliefs. Coherentism is like foundationalism in two important ways. First, both are internalist theories of justification. Second, neither theory explicitly employs (or implicitly assumes) any notion of significance.

Reliabilists hold that a belief is justified just in case it is produced by a reliable belief-forming mechanism. For example, visual beliefs are justified because our visual mechanism reliably produces accurate beliefs when applied to middle-sized objects. This raises the question of how to identify and individuate belief-forming mechanisms. The generality problem, probably the most serious objection to reliabilism, arises because there is more than one way to characterize the belief-forming mechanism that produced a particular belief. Some of these characterizations will denote a reliable process, whereas other characterizations will not. Without some way of deciding which of these processes to count as *the* one that produced the belief, the reliabilist runs the risk of having to say that such a belief is both justified (because it was produced by a reliable mechanism) and unjustified (because it was produced by an unreliable mechanism). And that's absurd (Goldman 1979, Feldman 1985). Unlike the previous two theories, reliabilism is an externalist theory of justification (since the justificatory status of a belief depends on the reliability of a belief-forming process, which is not a property entirely internal to the subject). But like foundationalism and coherentism, reliabilism does not employ a notion of significance. All three theories tell us which beliefs are justified; they do not tell us which beliefs are justified *and significant*.

2. Significance and standard analytic epistemology

Standard analytic epistemology focuses on notions that are supposed to be able to guide our cognitive efforts. But as we have made clear, the three traditional theories

of justification are not particularly concerned with the costs and benefits of reasoning. To a newcomer to epistemology, this might seem surprising. After all, we are finite creatures; yet, there are an infinite number of possible problem-tasks we could tackle at any particular time. Of course, most of these problem tasks ('What's 297 + 297?') never pop into our minds. However, we do occasionally find ourselves explicitly trying to refocus our scarce cognitive resources away from certain issues and toward others ('Gosh, I wish I'd stop humming that catchy but annoying tune and think about what to have for dinner' or 'Gee, I should stop thinking about why my spouse came home so late last night and focus on driving the car'). The effort to redirect our cognitive attention is an everyday event. Let's focus our discussion on a vivid example. Suppose Sue is watching a TV gameshow quiz on the same day she tested positive for breast cancer. A theory of justification can tell Sue whether she has arrived at a justified belief about a gameshow quiz question and whether she has arrived at a justified belief about her chances of having breast cancer given the positive result. But the traditional theories of justification offer no guidance about which of these issues to tackle (if she can only think about one of them) or about which is more significant.

Now, the defenders of traditional theories of justification will insist quite correctly that making these sorts of judgements is not what their theories are meant to do. Their theories provide an account of what distinguishes justified beliefs from unjustified beliefs—that is it. Further, the defenders of traditional theories can reasonably argue that the question of what issues are most significant for a reasoner to tackle are the jurisdiction of some other area of philosophy—perhaps action theory or moral psychology (or moral theory more generally).

The fact that traditional epistemological theories ignore issues of significance is not, by itself, a serious complaint against such theories. However, it does raise at least the possibility that contemporary epistemology unnaturally sunders belief from action. In order to explore this possibility, let's recognize the following principle: (P) While reason-guiding epistemological theories need not explicitly raise issues of significance, they must fit smoothly with our best views about and models of significance. This principle sets an indirect constraint on theories. Such constraints have been important in the history of science. For example, the phlogiston theory of chemistry held that burning a substance should lead to loss of weight—but there was a problem: metal gained weight after roasting. At first, caloric theorists could argue that caloric in metal had 'negative weight'. But phlogiston theory ran into serious problems with the success of Newtonian mechanics, where anything with mass could not have negative weight (Kuhn 1970: 70–72). The phlogiston theory could not easily survive the success of Newton's theory. We want to consider whether contemporary views about justification can smoothly and easily *supplement* their theories with accounts of significance. In order to explore this issue, let's turn to one particular way to think about the costs and benefits of reasoning.

3. A signal detection theory primer: one way to reason about significant issues

Reliability in our judgements is a good thing, but there is more to reasoning well than reliability. Coming to only true judgements is easy, if all one wants to do is come to true judgements. (For example, one could make a lot of true judgements about what one is not: I am not the number 1; I am not the number 2...) We want to make reliable judgements about significant matters. One way to represent this is with

Signal Detection Theory (SDT), widely used in sensory and perceptual psychology. SDT provides one computationally tractable way to represent an optimal pattern of correct and incorrect judgements given utilities assigned to those judgements. We will now look at the basics of SDT, but will not provide a formal and complete presentation of it here. An explication of the central features of SDT will be sufficient to illustrate the virtues of this approach and, in particular, spotlight its advantages.

Suppose that one had some evidence for thinking that some target category was present. To make this simple, we can think about a smoke alarm (although, as we shall make it clear, SDT has a wide range of real-world applications to human reason and decision making). The smoke alarm must employ two type of processes: a discrimination process and a decision process. A *discrimination process* assesses the extent to which evidence present in the observation favours the existence of one kind of stimulus category rather than another. So, for example, we can assume that the smoke detector detects the amount of smoke in the atmosphere. We can represent an idealized version of the situation facing the smoke detector (or any decision-maker) with the graph in figure 1. We have two bell curves: the horizontal x-axis represents the amount of smoke detected (i.e. the amount of evidence for thinking there is a fire) while the vertical y-axis represents the frequency with which the represented events occur. Let us suppose that the bell curve on the left represents the episodes in which the detector detects smoke but there is no fire (negative condition). The bell curve on the right represents the episodes in which the detector detects smoke and there is a fire (positive condition).

Our ability to discriminate these two conditions is depicted by their overlapping characteristics: the more they overlap, the more confusable are the conditions. A *decision process* involves deciding when the evidence is sufficient to judge that the target category (e.g. a fire) is present. The point at which the evidence is strong enough is called the *decision criterion* or *decision threshold*. The smoke detector's 'decision' criterion is the point (along the x-axis) at which the ambient atmosphere contains just enough smoke that the detector alarm gets activated. This decision may

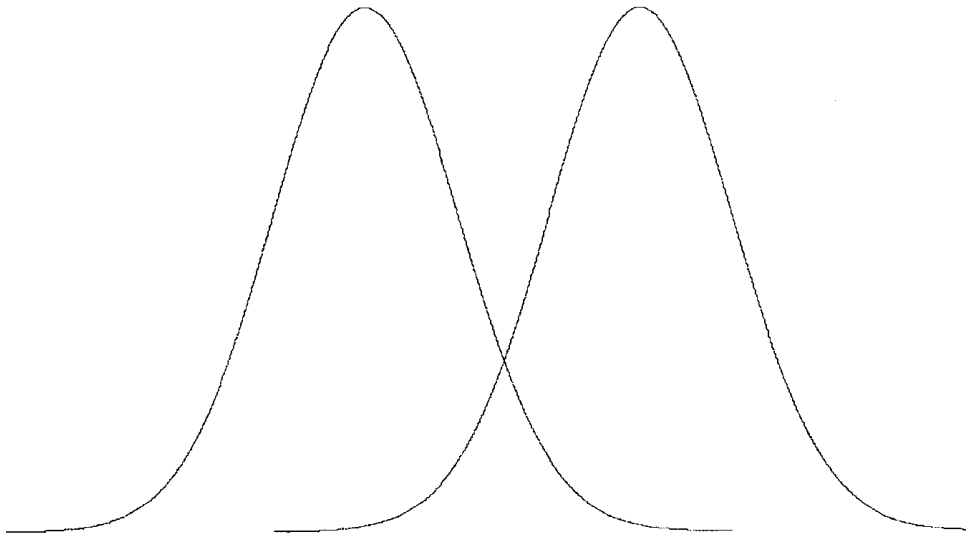

Figure 1. Representing the negative and positive conditions.

come early on in the accumulation of evidence, reflecting a lenient decision criterion, or much later, reflecting a strict or conservative criterion. SDT provides the machinery to find the optimum decision criterion.

According to SDT, there are two kinds of errors: false alarms (or false positives) and false negatives. There are also two kinds of correct judgements: true positives and true negatives. Here are four possibilities:

- H = hit = True Positive: true judgement that the target is present;
- M = miss = False Negative: false judgement that the target is not present (and it is);
- FA = false alarm = False Positive: false judgement that the target is present (and it is not);
- CR = correct rejection = True Negative: true judgement that the target is not present.

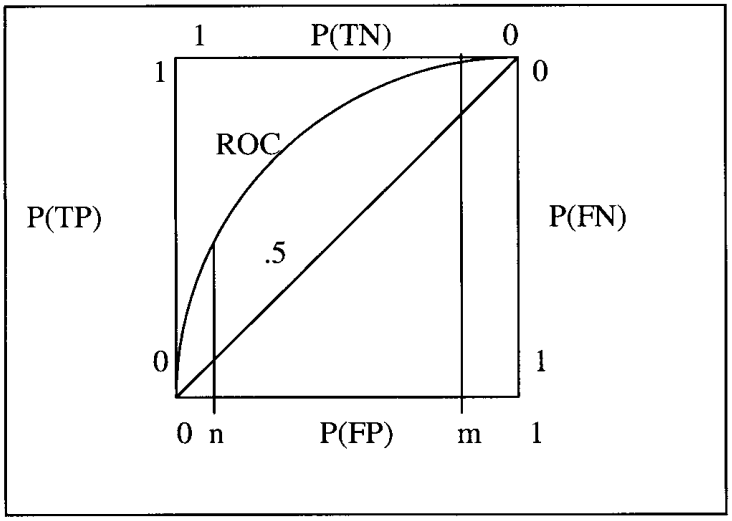

Figure 2. An idealized representation of the fire alarm's ROC.

Figure 2 is an idealized representation of a fire alarm's relative operating characteristic (ROC). It plots the relationship between the proportion of true positives (P[TP]) and the proportion of false positives (P[FP]) as the decision criterion varies. The area underneath the ROC curve is a measure of the fire alarm's accuracy. (If the 45 degree diagonal was the ROC, this would mean that the fire alarm was performing by chance.) One way to look at this measure of accuracy is in terms of the two characteristic normal distributions in figure 1. Insofar as the two curves share less area (i.e. are more separated), the fire alarm is more accurate (and so the ROC is pushed out farther from the diagonal). Perfect accuracy is, of course, rare. So, we might as well get used to handling inevitable error. SDT supplies some useful tools for helping us successfully navigate inevitable error. Note that the ROC makes clear the complementary relationships that hold between TP and FN on the one hand, and TN and FP on the other.

We can see how the ROC works by considering two possible decision criteria (at m and n in figure 2). Suppose the decision-maker sets a lenient decision criterion—one well to the right on the x-axis (at m) on graph 2. This criterion would result in the

decision-maker catching the vast majority of the true positives (and so would obviously have very few false negatives). But this would also mean that the decision-maker would have a high proportion of false positives (compared to true negatives). In the case of the fire alarm, an especially lenient setting would capture all of the possible hits—definitely going off when there is a fire in the house. However, it would also produce many false alarms, getting set off by a cigarette being smoked anywhere in the house, residual smoke from the fireplace or a slightly ambitious toaster. (In figure 1, this would involve setting the decision criterion toward the left side of the x-axis.) Setting the decision criterion here amounts to treating the task as though (for whatever reason) there is little cost in committing false alarms (saying that the target property is present when it is not). On the other hand, the decision-maker might set a very strict decision criterion—one well to the left on the x axis (at n) in figure 2. In this situation, the decision-maker is inclined to withhold a positive judgement, missing opportunities to detect the target when it is present. So, one might adjust the setting on the smoke detector to make it less sensitive to smoke. While it will not go off for cigarettes or toasters (true negative), it might miss real threats, going off only after a serious fire has filled the house with smoke. (In figure 1, this would involve setting the decision criterion toward the right side of the x-axis.) This conservative criterion treats the task as though (for whatever reason) there is little cost in missing a target property.[1]

Swets (1992) offers a number of real-life examples of decision-makers changing their decision criterion. Here is a particularly dramatic one:

> Recall that a few years ago the USS *Stark* suffered extensive damage when it failed to react until actually attacked by an enemy aircraft, and that more recently the USS *Vincennes* reacted incorrectly as if it were under attack and shot down a commercial airliner; we can surmise that in the interim the navy criterion for deciding that an attack is imminent shifted from strict to lenient. (1992: 524).

One handy feature of SDT is that it yields a relatively simple formula for finding the optimal decision criteria. The formula requires information about: (a) the prior probabilities of the alternative conditions; (b) the net benefits of a negative judgement, i.e. benefit(TN)–cost(FP); and (c) the net benefits of a positive judgement, i.e. benefit (TP)–cost(FN). So, for the fire alarm, the optimal decision criterion is the point on the ROC whose tangent has a slope that equals:

$$\frac{\text{P(no fire)}}{\text{P(fire)}} \times \frac{\text{net benefit of a 'no fire' judgement}}{\text{net benefit of a 'fire' judgement}}$$

Perhaps the most intuitive way to understand this is that larger numerators tend to make the optimal decision criterion more strict (leftwards in figure 2, rightwards in figure 1); larger denominators tend to make the optimal decision criterion more lenient (rightwards in figure 2; leftwards in figure 1).

- *P(no fire)*: if this is large, that means that in figure 1, the left bell curve is considerably taller than the right bell curve (i.e. smoke most often means there is no fire). Other things being equal, in order to avoid lots of false alarms, the decision criterion in figure 1 should be pushed rightwards.
- *Net benefit of a 'no fire' judgement*: suppose the fire alarm doesn't sound (makes a 'no fire' judgement). If it is right not to sound (i.e. there is no fire), this brings a benefit; but if it is wrong not to sound (i.e. there is a missed fire), this brings a cost. The cost of a missed fire (i.e. a false positive) is huge compared to the

benefit of a true negative. So other things being equal, this should push the decision criterion in figure 1 leftwards—we want to make sure we do not miss the real fires.

- *P(fire)*: if this is large, that means that in figure 1, the right bell curve is considerably taller than the left bell curve (i.e. smoke most often means fire). Since false alarms will not occur very often, the decision criterion in figure 1 should be pushed leftwards.
- *Net benefit of a 'fire' judgement*: suppose the fire alarm sounds. If it is right to sound (true positive), this brings a large benefit. If it is wrong to sound (false negative), this brings a cost (mostly in terms of annoyance). Since the benefit is considerably greater than the cost, other things being equal, this should push the decision criterion in figure 1 leftwards—we want to make sure we catch the real fires.

Determining the optimal decision criterion will demand intellectual (and sometimes physical) legwork. First, we must do the empirical legwork of finding out the prior probability (or base rate) of the presence of the target property (breast cancer, mental illness, success in a job, etc.). Second, we must explicitly identify our priorities by reflecting upon the value we place on the benefits and costs of TN, TP, FP and FN in specific cases.

Let us now leave behind the smoke detector, which we fear is going to confirm people's judgement that SDT can be used only for a narrow range of prediction problems. Let us consider the detection of breast cancer. False positives (i.e. misses) can be deadly. False negatives range from anxiety provoking to the merely irritating; that is, missing cancer is much worse than saying that you have it when you don't. So this would tend to lead us think that the decision criterion should be lenient. But it is important to recognize that accuracy has everything to do with the proportion of positive and negative conditions in the population. If we are attempting to identify women with breast cancer in a referral population in which most of the women have a malignancy, an especially lenient criterion is appropriate. In a population of non-symptomatic women in which few will have a malignancy, however, a stricter criterion is appropriate for screening. It is important to remember that arriving at our decision criterion requires that we consider the prior probabilities of the two conditions as well as the attractive or aversive values of the four possible reasoning outcomes.

Error is the unwelcome companion of all judgements. So, we need to identify the sources of error that can be regulated and steer them to our advantage. The quickest route to maximizing utility is to use a reliable reasoning strategy of low-relative cost, and apply it to problems that are significant. We may make the extra-statistical decision that rule X addresses an especially significant property, and makes such applications top-priority. However, we do not want predictive accuracy *simpliciter*. What smart shoppers want is the greatest accuracy given their global priorities, and in order to get it, they must manage the decision-maker's proneness to error. SDT has shown itself useful in a wide range or problems. As long as the problems can be tackled with a discrimination and decision structure, SDT can be usefully applied. (SDT is not useful in population genetics, for example.) Below is a truncated list of successful applications of SDT:

(1) the accuracy of visual, auditory, olfactory, gustatory and tactile judgements;
(2) the accuracy of memory;

(3) the success of students in undergraduate, graduate, law school, and medical school;
(4) the success of the job applicants;
(5) the likelihood of violent behaviour by candidates for involuntary hospitalization;
(6) the presence of a disease (e.g. breast cancer, prostate cancer, psychosis, etc.) and
(7) the success of the course of treatment; and
(8) the success of a social programme.

These are, by anyone's standards, very significant judgements facing us. Although we extol the virtues of SDT in making accurate predictions about such complex and significant problems, we do not want to imply that these are virtues *unique* to SDT. Many cognitive prosthetics can assist crude subjective judgement. As Swets *et al.* (2000) point out, we can use as the criterion of predictive validity the signal detection statistic of the area under received operating curves (ROCs), correlation coefficients or accuracy of categorization. There are no differences between these other methods and SDT in terms of the in the direction or the implications of the results.

The task of employing the machinery of SDT to assign optimum decision criterion values may at first appear foreboding, but remember two things. First, the investment of effort in filling out this formalism would seem a low price to pay for improved performance on significant problems. Second, we already make these sorts of calculations on the fly, and so for significant problems we might as well make them competently. Now, what makes a problem significant? This is thorny terrain, but it must be crossed. Let us take a few steps, and these will be a few steps more than most traditional epistemologists have taken.

4. A schematic account of general significance

In light of toenail length examples, people readily agree that an epistemology ought to guide us to significant truths, or at least not endorse principles inimical to their discovery. That we preferentially desire significant truths, then, is beyond practical dispute. The topic of epistemic significance, however, has captured little attention, and even less sustained treatment. However, Kitcher (2001) has recently developed an account of epistemic significance designed to characterize relations of importance in science, and it provides a basis for our discussion. Kitcher begins with the observation that. 'The aim of science is not to discover any old truth but to discover significant truths' (2001: 87). According to his account, scientific aims are ordered by interests that individuals have in addressing, and ultimately solving, general and deep theoretical questions in science, questions such as 'How do organisms develop?' or 'How did our universe form?' Scientists address these deep questions by investigating a wide range of more specific inquiries. The significance of these more specific inquiries is a function of how they are related to the more fundamental questions. In particular, these significance relations are best depicted with a graph in which the significance of specific (often experimental) findings flow from their contribution to deep and fundamental theoretical questions.

Kitcher's account is explicitly sensitive to resource allocation concerns; indeed, the question of epistemic significance typically arises when asking whether a hypothetical payoff is sufficiently valuable (that is, significant) to warrant a given investment of resources to a problem. In science, the ultimate allocation decision is based on a

judgement about the reliability of the methods and instruments of a research programme, the talent and commitment of the personnel, the promise of the research programme and the significance of the basic issue it addresses.

On Kitcher's view, the tracking of overall scientific significance by specific decisions depends on the availability of a 'well-ordered science' (2001: 117–135). Our account of epistemic significance, however, concerns not just the epistemology of science proper, but epistemology in general. Much of our knowledge is acquired and deployed without arcane skill, careful study or appreciation of theoretical detail—ordinary perceptual knowledge, for example. A complete epistemology should offer guidance in both scientific and lay life.

A general account of epistemic significance is bound to rest, ultimately, in judgements about what conduces to human well-being. If, in keeping with naturalism, we view humans as part of the natural order, then the conditions that contribute to human well-being are open to scientific investigation. It is an empirical question whether the conditions for human well-being are too variable to be studied systematically and generalized about scientifically. There is, however, reason for optimism. While the conditions of human well-being may be tied to norms that differ across individuals and across cultures, the difficulties of generalization across individual and group have been surmountable challenges in other sciences, such as developmental and evolutionary biology. At the core, the conditions of human well-being concern health, deep social attachments, personal security and the pursuit of significant projects. Epistemic norms that most deeply frustrate these goals are more likely than others to provide poor payoffs.

5. Standard analytic epistemology, again

Now let us pull together the various strands of argument. Contemporary analytic epistemology focuses most of its attention on theories of knowledge and justification. One reason for this focus is that such notions are supposed to be *reason-guiding*, but we pointed out that contemporary theories of justification do not explicitly mention or implicitly employ some notion of epistemic significance. Such a notion seems essential if an epistemological theory is going to have a chance to successfully guide our reasoning. So we have described one way to supplement theories of justification with some systematic view about the nature of epistemic significance—with an account that gives a central place to Signal Detection Theory. We now want to explore how well the standard accounts of justification are likely to fit with the SDT-based account of epistemic significance, and (more generally) with any account of epistemic significance.

5.1. *The unreliability of our epistemic intuitions*

Internalist views of justification (foundationalism, coherentism) hold that justifiers must be accessible to the subject—they are available (perhaps directly) to the subject's conscious awareness. However, even if justifiers are available to our awareness, it does not follow that we can readily identify them as belonging to justified beliefs rather than unjustified beliefs. For example, when deciding whether to believe p or not-p, a subject, after careful introspection, might come to believe that the justifiers belong to the belief that p, when in fact they belong to the belief that not-p. There is considerable evidence that people sometimes judge that the decisions recommended by decision prosthetics (like SDT) are mistaken and people sometimes

judge the beliefs recommended by reasoning prosthetics to be mistaken (see Bishop and Trout 2002). In other words, people's intuitive epistemic judgements do not accord with the best available models for reasoning and decision-making.

Any epistemological theory that pretends to be reason-guiding must face up to this systematic discrepancy between people's epistemic intuitions and the pronouncements of our best models. Using a lifeless algorithm to identify the presence of cancer may lead to beliefs and decisions that do not satisfy our well-considered intuitive normative judgements; using unaided judgement to decide whether to trade in or out of an equity may be very intuitively satisfying. In the long term, however, the algorithm rewards those who apply it ruthlessly, without weakening in the face of the subjective pull of intuition. When predictive accuracy means life or death, the beach house or the poor house, successful knowers (almost always) prefer the mechanical procedure. (For a discussion of the rare exceptions, see Bishop and Trout 2002.)

The fact that our epistemic intuitions systematically differ from the judgements of the best available reasoning and decision-making models raises a dilemma for internalist theories of justification. In any particular case of discrepancy, an internalist theory must hold either that our intuitive judgement about which belief is justified is correct (and that the model is incorrect) or that our intuitive judgement about which belief is justified is incorrect. If the internalist theory takes the first horn of the dilemma, then the internalist has defined a notion of justification that is not truth-conducive. (That is, there is some way of combining the subject's evidence so that her beliefs are systematically more reliable in a non-sceptical world than if she were to consistently adopt beliefs that are 'justified' according to the internalist.) If the internalist theory takes the second horn of the dilemma, then the internalist has defined a notion of justification that is systematically mis-identified by subjective introspection. In either case, the internalist must admit some kind of systematic error on the part of reasoners. What is worse is that this error is unnecessary—it could be overcome by adherence to the successful model.

Presumably, any epistemological theorist who takes seriously the goal of guiding people's reason is going to find this kind of unnecessary systematic error deeply problematic. Uncorrected, it will lead people to fewer truths and worse decisions. The only way to overcome these unnecessary errors (and their deleterious effects) is to explicitly make reference to considerations that are external to the reasoner. That is, the internalist theory is going to have to be supplemented with some procedure for overcoming reasoners' unnecessary and systematic errors. We are not contending that some of the internalist's *justifiers* must be external to the subject (which would mean the end of internalism). The internalist's justifiers might all be internal to the subject. However, in order to be reason-guiding, the internalist is going to need to make essential appeal to external considerations.

Suppose this is right—the internalist is going to have to supplement her theory with a corrective theory that appeals to considerations external to reasoners. So what? Is this really a problem for internalism? After all, the central internalist thesis (all justifiers are internal to the reasoner) survives. But we think this is a problem. A fundamental assumption of contemporary epistemology is that epistemology should be able to guide reasoning. One of the primary motivations for internalism is that reasoners can only be expected to consider those factors that are *readily available* to them—the justifiers for S's beliefs must be readily available (or knowable) by S. However, as Goldman has pointed out, this constraint does not 'rationalize' internalism (1999: 288). After all, there are plenty of external facts that are more

easily available than many internal ones. It is probably easier for you to figure out whether you are currently sharing a room with an elephant than it is for you to recollect what you ate for dinner last night, let alone what spices it contained. However, once the internalist admits that in guiding reasoning, she must advert to facts external to the reasoner, this motivation for internalism is lost. After all, in offering guidance to reasoners, both the internalist and the externalist are going to be essentially adverting to facts external to the reasoner. The fact that the internalist's justifiers are all internal to the reasoner and the externalist's are not becomes a mere conceptual nicety. On both views, the normative, epistemic considerations that guide reasoning are not entirely internal to the subject. Which of the available theories employ external factors that are the most readily available to the reasoner (and so most easily employed by the reasoner)? Is the theory that is most easily used by reasoners an internalist theory or an externalist theory? These are thoroughly open questions.

5.2. *Access worries*

Internalism holds that justifiers are internal to a subject. However, when people make decisions, they are typically unaware of where precisely they have placed their decision criterion. It takes SDT or a similar instrument to determine precisely where an individual's decision criterion is located. Consider the accuracy of an individual medical professional who diagnoses breast cancer. Surveying the ultimate outcomes of hundreds of his or her diagnoses based on mammograms, we can identify her decision criterion poised between two distributions. The diagnostician will discover that she had a decision criterion of a specific value (or location), whereas prior to this graphing of outcomes she, at best, had only an impression of its position within a broad range, such as 'lenient'. One thing this means is that our power to introspect our rationale for a particular decision is cognitively limited; if we have to look at a graph to see when a situation supplied enough evidence for us to make a decision, or how normal variation influences us, then we do not have complete epistemic authority over the processes that lead to that decision. After all, if you have access to an internal justifier, who needs a cognitive prosthetic like a graph? Well, apparently, we do.

Moreover, awareness of this limitation does not seem to make an appreciable difference to judgement by traditional unaided methods, whether acquainted with SDT or not. Thus, internalist approaches to epistemic reform must concede that the internal justifier of judgement is unreliable and (whether in principle inaccessible or not) knowing that it is unreliable is epistemically unhelpful without the assistance of cognitive prosthetics. Internalists may reply that only the grounds of justified belief need be accessible, not the grounds for action. The position of the decision criterion is a measure not of belief but of performance, and so its inaccessiblity leaves internalism untouched. The internalist may try this gambit, but it creates an alarmingly unprincipled account of justification. Recall that the virtue claimed for internalism was its normative, reason- and action-guiding properties; guiding reason is important mainly because reason guides action. It would be a peculiar theory of justification that declares its action-guiding function via justified belief, but then ducks the disconfirming products of poor action-guidance by renouncing the relation between belief and action.

5.3. *The potential ameliorative effect of a SDT-based view of significance*

There is a considerable literature that suggests that people's judgements tend to be overconfident.

> [A] large majority of the general public thinks that they are more intelligent, more fair-minded, less prejudiced, and more skilled behind the wheel of an automobile than the average person... A survey of one million high school seniors found that 70% thought they were above average in leadership ability, and only 2% thought they were below average. In terms of ability to get along with others, *all* students thought they were above average, 60% thought they were in the top 10%, and 25% thought they were in the top 1%! Lest one think that such inflated self-assessments occur only in the minds of callow high-school students, it should be pointed out that a survey of university professors found that 94% thought they were better at their jobs than their average colleague. (Gilovich 1991: 77)

The overconfidence bias goes far beyond our inflated self-assessments. For example, Fischhoff *et al*. (1977) asked subjects to indicate the most frequent cause of death in the USA, and to estimate their confidence that their choice was correct (in terms of 'odds'). When subjects set the odds of their answer's correctness at 100:1, they were correct only 75% of the time. Remarkably, even when they were so certain as to set the odds between 10 000:1 and 1 000 000:1, they were correct only between 85% and 90% of the time. It is important to note that the overconfidence effect is systematic (it is highly replicable and survives changes in task and setting) and directional (the effect is always in the direction of over rather than underconfidence).

Signal Detection Theory, it turns out, holds uncaptured power to regulate and correct such self-serving assessments of our skills and knowledge. It allows us to locate our decision criterion for some particular problem-set. By comparing it to the hypothetical decision criterion with the desired payoff characteristics, we can calibrate our judgements in the direction of performance that we actually desire, correcting our anecdotal and subjective (and inaccurate) impressions of our own accuracy.

6. Conclusion

Epistemology is a serious business, in part, because our beliefs so often play a decisive role in how we decide to intervene in the world. We contend that contemporary epistemology would do well to focus a bit more attention on how their normative theories of reasoning might connect up with normative theories about action. We have suggested the use of SDT (and other such methods) to capture our considered utilities. Problems that are epistemically significant at the moment might not be significant tomorrow, next year or even in ten seconds. But while what problems are significant can vary dramatically, the notion of significance is neither irretrievably subjective nor so flexible across time and contexts that it is intellectually intractable. The demands of human well-being impose a natural order on what problems can count as significant. These demands may stand as only outside constraints, but they are constraints nonetheless.

A mature epistemology should tell us how to reason effectively about issues, both important and trivial, that confront us daily. Such issues are often complex. Given the documented success of various cognitive aids for reasoning and decision-making, it is reasonable for a mature epistemology to seek out and (when possible) take advantage of these cognitive aids. These aids can help us to balance accuracy and significance; they can help us to formalize epistemic norms so that we can execute our practical goals with accuracy, efficiency and, ultimately, satisfaction. It is quite

likely that these epistemic norms will be, in a certain respect, hypothetical. They will not tell us what to do proper, but what to do relative to our sincerely held goals: if your goal is X, then (given the payoff schedule) you should do Y in attempting to achieve it.

This way of thinking about thought and action raises a sensitive issue. The action-guiding reach of epistemology has the potential to tackle practical and significant affairs: issues about crime, medical diagnosis, poverty, etc. However, epistemologists have spent a great amount of time, talent and energy developing accounts of justification that are not obviously connected to action (unless one counts the adoption of belief as an action). Nevertheless, it is common for philosophers to commend the benefits of seeking justification. If justification is supposed to be action-guiding simply because it prompts the mere act of belief, it is compatible with believing much and doing nothing else. However, as with any decision, there is a price to pay for attachment to an intellectual tradition that sunders justification from effective action. As far as we know, no one has even attempted to calculate the cost.

Note

1. We can determine the reliability of a strategy only for processes whose outcomes can be determined. If you are testing instruments that purport to identify criminals prone to recidivism, there is a state or condition (the person is caught doing the crime again) whose existence would provide evidence for or against the reliability of the rule. While it is possible that the person could commit the crime again without getting caught—and so the test instrument would not be as accurate as thought—the problem of undetected events is ubiquitous. For a philosophically illuminating use of SDT in the philosophy of mind, see Godfrey-Smith (1991).

References

Alston, W., 1989, *Epistemic Justification* (Ithaca, NY: Cornell University Press).

Bishop, M., and Trout, J. D., 2002, 50 years of successful predictive modeling should be enough: lessons for philosophy of science. *Philosophy of Science: PSA 2000 Symposium Papers*.

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

Feldman, R., 1985, Reliability and justification. *The Monist*, **68**: 159–174.

Fischhoff, B., Slovic, P., and Lichtenstein, S., 1977, Knowing with certainty: the appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, **3**: 552–564.

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

Godfrey-Smith, P. 1991, Signal, decision, action. *The Journal of Philosophy*, **91**: 709–722.

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

Goldman, A., 1999, Internalism exposed. *The Journal of Philosophy*, **96**: 271–293.

Kitcher, P., 2001, *Science, Truth, and Democracy* (New York: Oxford University Press).

Kuhn, T., 1970, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press).

Pollock, J., and Cruz, J., 1999, *Contemporary Theories of Knowledge*, 2nd edn (Totowa, NJ: Rowman & Littlefield).

Swets, J., 1992, The science of choosing the right decision threshold in high stakes diagnostics. *American Psychologist*, **47**, 522–532.

Swets, J., Dawes, R., and Monahan, J., 2000, Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, **1**: 1–26.