



# Strategic Reliabilism: A Naturalistic Approach to Epistemology

Michael A. Bishop  
*Florida State University*

J. D. Trout\*  
*Loyola University Chicago*

---

## Abstract

Strategic Reliabilism is a framework that yields relative epistemic evaluations of belief-producing cognitive processes. It is a theory of cognitive excellence, or more colloquially, a theory of reasoning excellence (where ‘reasoning’ is understood very broadly as any sort of cognitive process for coming to judgments or beliefs). First introduced in our book, *Epistemology and the Psychology of Human Judgment* (henceforth *EPHJ*), the basic idea behind SR is that epistemically excellent reasoning is efficient reasoning that leads in a robustly reliable fashion to significant, true beliefs. It differs from most contemporary epistemological theories in two ways. First, it is not a theory of justification or knowledge – a theory of epistemically worthy *belief*. Strategic Reliabilism is a theory of epistemically worthy ways of *forming* beliefs. And second, Strategic Reliabilism does not attempt to account for an epistemological property that is assumed to be faithfully reflected in the epistemic judgments and intuitions of philosophers. If SR makes recommendations that accord with our reflective epistemic judgments and intuitions, great. If not, then so much the worse for our reflective epistemic judgments and intuitions.

---

## 1. Strategic Reliabilism

Strategic Reliabilism (SR) has three parts:

1. The core theory, which articulates the standards according to which cognitive processes are evaluated.
2. The theory of the proper acquisition and revision of cognitive processes.
3. The applied theory, which recommends specific reasoning strategies for particular domains.

Most of our attention will be devoted to the core of the theory. Strategic Reliabilism evaluates cognitive processes in terms of the extent to which they are a suitable combination of (a) robustly reliable; (b) appropriate to the reasoner’s resources; and (c) geared toward producing beliefs about topics that are significant for the reasoner.

## 1.1. ROBUST RELIABILITY

Suppose Susan faces a reasoning problem. We can think of a reasoning problem as a question with a range of possible answers; e.g., will this applicant succeed in graduate school (where there is some clear criterion for 'success')? There are many possible ways Susan might reason through this problem. By closely observing Susan's judgments and what she says about why she makes her judgments, we can describe in rough terms what considerations she takes to be relevant and how she weighs them. For example, she might give significant weight to the letters of recommendation and almost no weight to the quality of the applicant's undergraduate institution. Her colleague might weigh these lines of evidence quite differently. Let's call these different ways of reasoning about the problem, *reasoning strategies*. A reasoning strategy is a repeatable pattern of reasoning which is instantiated by a cognitive process. The basic idea behind SR is that the various reasoning strategies Susan might have used are ranked according to their reliability. Susan's reasoning is epistemically better to the extent she used a more reliable reasoning strategy.

The reliability score of a cognitive process is its truth ratio (i.e., its ratio of true to total judgments) on its *expected range of problems*. Given a reasoner's environment, there is a certain distribution of problems she can expect to face. And given a reasoner's cognitive dispositions, each of her cognitive processes will apply to a subset of those problems. These problems are the cognitive process's (or reasoning strategy's) expected range of problems. We can test the strategy on a representative sample of problems on its expected range to find its observed reliability score in that range. This will approximate the reasoning strategy's real reliability score for that person in that environment. Applying this to Susan's situation, we can take the various reasoning strategies Susan might have used to make admissions decisions, apply them to a representative set of applicants, and come up with an observed reliability score for each of the strategies. If we do this well, the observed scores will closely approximate the real scores. And so we can rank the epistemic quality of these various reasoning strategies in terms of their reliability.

Cognitive excellence involves *robust* reliability. A robust cognitive process is one that is reliable across a wide range of environments. Consider the visual system. It is usually quite reliable. But color perception under artificial light can be unreliable. So we can partition the visual system's range into those problems dealing with color perception under artificial light and all the other problems. This is a *discriminable* partition because it is based on properties that can be detected by a reasoner prior to her using the strategy. A cognitive process is more robust to the extent its reliability score is *consistently reliable* across various discriminable partitions of its range. So robustness is a matter of scope and consistency. (1) Scope: A process is more robust to the extent it can be applied to a larger set of

problems. (2) Consistency: A process is more consistent to the extent it is reliable across the various discriminable partitions in its range.

There are often trade-offs to be made between scope and consistency. Take a cognitive process that is somewhat inconsistent – it is reliable for some discriminable problem-types in its range but not for others. For example, our color perception is generally reliable, but is significantly less reliable in artificial light. Strategic Reliabilism might recommend that we not trust our color perception of objects in artificial light. This reduces the scope of the process (by limiting its appropriate range) and bolsters its consistency (by eliminating from its appropriate range some discriminable partitions in which the process is unreliable). It will often be possible to raise a reasoning strategy's reliability by reducing its range. So SR must make some non-trivial trade-off judgments: At what point is greater scope (and less reliability) preferable to greater consistency (and more reliability)? Resource considerations are vital in making these trade-offs. That's because as we raise the reliability of a reasoning strategy (or a cognitive process generally) by reducing its range, we typically make the rule harder to use by making the detection of its appropriate range more complicated.

## 1.2. THE COSTS AND BENEFITS OF EXCELLENT REASONING

Strategic Reliabilism commits us to the idea that the excellent reasoner will *efficiently* use her limited cognitive resources. The notion of efficiency requires a conception of the costs and benefits of reasoning. The costs are the resources we expend in reasoning, and the benefits are believing significant truths. Since we don't know how to measure these items, we suggest using measurable proxies that track reasonably well the costs and benefits of reasoning. When we suspect that the proxy does not track the costs or benefits of reasoning, we can take this into account in applying the theory. The proxy for reasoning costs is time – the time it takes to implement a reasoning strategy (start-up costs) or to use that strategy to solve a problem (execution costs). The proxy for reasoning benefits is true belief. The benefit of good reasoning is not true belief, per se, but true belief about significant matters. The person who reasons reliably about significant matters is a better reasoner than the person who reasons even more reliably about useless, insignificant matters. There are many ways to characterize significance. We defend a view that assumes there are objective reasons for us to tackle some problems rather than others (*EPHJ* 95–9). Some might prefer a subjective or pragmatic view about what makes a reasoning problem significant. We need not take a firm view of the matter here.

## 1.3. STRATEGIC RELIABILISM IN ACTION

When conjoined with well-established psychological findings, SR has the resources to offer a surprising amount of specific, useful advice about how

people ought to reason. In chapters 3 and 9 of *EPHJ*, we describe simple heuristics we can use to avoid overconfidence, to improve how we reason about causal matters, and to improve how we reason about the results of diagnostic tests (e.g., tests for disease, pregnancy, drug use). This advice is non-obvious and practically significant. Rather than review that advice, we will offer new advice about how to avoid the planning fallacy – the ubiquitous tendency to underestimate the resources it will take to complete a project. Famous examples of the planning fallacy abound. Construction of the Sydney Opera House went an astonishing 10 years and \$95 million beyond projections (Eikeland). The ‘Channel Tunnel’ or Chunnel was ‘only’ a year late but 80% (\$4.5 billion) over budget (Flyvbjerg). And the Denver airport was 16 months late, and \$2.8 billion over budget (de Neufville, ‘Baggage System at Denver’; ‘Airport Privatization’). These telling examples are dramatic (see Trout, *Empathy Gap*), but the planning fallacy is an everyday occurrence. Be honest: When moving, how many times have you fallen victim to the planning fallacy and been forced to work through the night, packing and cleaning?

The planning fallacy occurs because we adopt an ‘internal’ approach to predicting how long it will take us to complete a project. We begin with a narrative that breaks down the project into component steps, we estimate how long each step will take, perhaps including some ‘slippage factor’ time, and then we add it all up. This narrative does not include any specific mishaps – we don’t have enough boxes, our friends can’t help with the packing, the basement floods – because the chances of any one of these things going wrong is low. We might recognize that the chance that *something* will go wrong is reasonably high, and so we add the slippage factor. But it’s not typically enough. An obvious suggestion for overcoming the planning fallacy is to train yourself to add more slippage time. But how much? There is another solution that is more radical and more reliable: Avoid the internal approach altogether, and don’t start by imagining a scenario. Instead, adopt an ‘external’ approach that begins by considering your past history with relevantly similar tasks: How long has it taken you to move in the past? Start there and revise by comparing the current task to past ones (Kahneman and Tversky).

Let’s consider how the external approach works with an example that will be familiar to many philosophers. You are asked to contribute an article, due in 6 months, to a volume edited by a distinguished colleague. Can you do it? You might adopt the internal approach and think, ‘A 20-page paper on the endurance-perdurance distinction, how hard could it be? I’m a fast writer, and my teaching load is fairly light this semester’. The external approach recommends you start by looking at your history with relevantly similar projects. Perhaps you have records of how long it has taken you to write articles of this sort in the past. Or you might look at your CV and figure out how many articles you write per year. Start with this ‘external’ preliminary projection. And then adjust it according

to whether there is anything unusual about the article solicited that would make it unrepresentative of those you're comparing it to. In making this adjustment, we must fight the temptation to allow overly optimistic, scenario-based internal factors in through the back door. For example, we might make the external prediction considerably more optimistic on the grounds that past inefficiencies no longer exist (the kids were young, the teaching load was heavier, I was in the middle of a nasty divorce, etc.). By imagining a scenario in which we work on the project in the absence of these factors, we are likely to ignore past efficiencies that have been lost and novel inefficiencies that cannot be predicted.

## 2. Some Objections to Strategic Reliabilism

We will now consider six objections to SR.

### 2.1. STRATEGIC RELIABILISM AND WISHFUL THINKING

Some have taken SR's reliance on significance to imply a case for wishful thinking: 'If good reasoning must address itself to positively significant problems, as the authors claim, then Hume's brilliant essay on the immortality of the soul must be bad reasoning' (Goldman, 'Review'). Strategic Reliabilism does not have this result. Significance is a property of reasoning problems. Strategic Reliabilism tells us that reasoning excellence is reliable reasoning about significant problems. It does not say that reasoning excellence is reasoning that results in beliefs that make us happy or contribute to our well-being. The only role significance plays in SR is in *directing attention and resources* to certain problems and away from others. Once a problem is engaged, reasoning excellence requires only reasoning that is as robustly reliable as possible, given the reasoner's resources.

### 2.2. STRATEGIC RELIABILISM AND 'PURE' EPISTEMOLOGY

Strategic Reliabilism is a theory that yields epistemic evaluations of how people reason (where 'reason' is understood broadly) and it prescribes one way of reasoning that is epistemically best (or more than one if there is a tie). Now let's add this assumption: The epistemic quality of a belief is a direct function of the epistemic quality of the reasoning that produced it. With this plausible assumption, SR can evaluate and recommend beliefs – these normative judgments will 'piggyback' on the epistemic evaluations and recommendations SR makes about a person's reasoning.

What is the relationship between the recommendations of SR – particularly the recommendations about what one epistemically ought to believe – and theories of epistemically justified belief? Some of our critics have argued that SR does not conflict with theories of justification. One line of argument contends that SR does not concern itself with the 'purely

epistemic ought' but only with the 'ought simpliciter'. And so SR and theories of justification are not competitors.<sup>1</sup> But this objection relies on a faulty reading of SR. Strategic Reliabilism does not concern itself with how we ought (all things considered) to reason. Simplifying somewhat, SR says that one ought to reason reliably about significant problems. This is an epistemic 'ought' and it can conflict with moral or prudential 'oughts'. For example, reasoning reliably is not always in our best interest (see section 2.6 below).

Along similar lines, Ram Neta has suggested that SR is a 'mongrel' epistemological theory because it allows pragmatic considerations like significance and resource limitations to play a role in determining what counts as the best reasoning for a person (333). Neta argues that an 'Unsullied Reliabilism' (UR, 'epistemic excellence consists solely in the use of reliable reasoning strategies') is better supported by the evidence. Neta's argument consists of two premises and a conclusion:

1. The empirical evidence in support of SR supports UR just as well (334).
2. The prior probability of UR is higher than the prior probability of SR (334).
3. Therefore, the posterior probability of UR will be higher than that of SR.

The first premise is false. Neta recognizes that a major part of the empirical evidence that supports SR is Ameliorative Psychology: 'those branches of cognitive psychology . . . that study good reasoning' (328). One of our primary arguments for SR is that it grounds and explains the successful normative, epistemic judgments of Ameliorative Psychology. On this score, SR is superior to UR for two reasons. First, Ameliorative Psychology recommends reasoning strategies that tend to focus on significant problems – e.g., diagnosing illness and predicting violence (*EPHJ* 57–8). Second, the recommendations of Ameliorative Psychology are resource sensitive – *ceteris paribus*, more tractable reasoning strategies are evaluated more positively than less tractable ones (*EPHJ* 56–7). Unsullied Reliabilism is blind to these real patterns in the evidence. Only a theory that favors more efficient reasoning strategies that tackle significant problems – a theory like SR – can account for these features of Ameliorative Psychology. And so the evidence supports SR much better than it does Neta's UR.

Neta supports the second premise with a hypothetical case of a brilliant and highly reliable analytic epistemologist. Intuitively, she reasons in ways that are epistemically excellent even though the problems she considers are 'not . . . significant' (332–3). Strategic Reliabilism is not committed to this. Significance is a property of a problem for a reasoner (at a time); and the problems of analytic epistemology are certainly significant for lots of people (on pragmatic and professional grounds at the very least). But putting this objection aside, it's not obvious that SR does worse than UR on intuitive grounds. Consider a series of cases, beginning with a reasoner obsessively focused on an obviously (almost) insignificant problem (e.g., the distance between her shoelaces) and ending with a reasoner focused on a highly

significant problem (e.g., how to avert an impending nuclear war). All the reasoners in the series reason equally reliably. Do we really want to say that the shoelace-obsessed reasoner is as epistemically excellent a reasoner as the others, even if he is constantly falling down stairs, walking in front of cars, and forgetting to eat or bathe for days at a time? No doubt, some epistemologists have Neta's intuitions in this case. We don't. But is that really the best we can do – battle to an intuitive standoff? We think not. On our view, any reasonable epistemological theory should account for the contingent, empirical, positive correlation between good reasoning and good outcomes. We call this the Aristotelian Principle: In the long run, those who reason better tend to have superior pragmatic results. Neta's UR gives us truth, and truth is grand, but most truths aren't useful. Purity is vain if it requires that we judge the shoe-lace obsessed subject to be as good a reasoner as Darwin and its guidance is impotent insofar as it delivers resource-independent recommendations that no human can implement.

### 2.3. THE CONTOURS OF STRATEGIC RELIABILISM

In very restricted domains, SR can be applied to test various computer models against each other on a wide range of real-world problems to determine how reliable they are. But in real-life situations, the sort of evaluations SR calls for seem practically impossible. Jonathan Weinberg (personal communication) has articulated this point nicely (although as far as we know, he is not committed to it): Strategic Reliabilism cannot give an account of the full breadth and variety of our epistemic practices; there are only a very select and limited set of cognitive tasks for which we can even begin to perform the kind of cost-benefit calculations that Strategic Reliabilism requires.

The objection is that SR fails to meet some standard that epistemological theories ought to meet and that (presumably) plenty of other available epistemological theories do meet. What might that standard be? If the demand is that a theory provide us with a practically useful way to 'account [for] the full breadth and variety of our epistemic practices', no extant theory meets it. That's just not a reasonable expectation. Perhaps the point is that for a wide range of real-world beliefs (or reasoning that led to those beliefs), a theory should provide a practically manageable way to assess those beliefs (or those instances of reasoning). But on that standard, it's not obvious why SR does worse than reliabilist theories of justification – or for that matter, internalist theories of justification. Goldman ('Internalism Exposed') and Kornblith (*Knowledge and Its Place*), for example, have argued that whether a person's belief is coherent with the rest of her beliefs is an extraordinarily difficult matter to assess. So without a lot more argument, we see no reason to agree that SR deserves criticism for being more difficult to apply to real-world cases than most other epistemological theories.

But we want an epistemological theory to do more than simply ‘account’ for our *actual* epistemological practices. We want a theory that *recommends better* epistemological practices. Strategic Reliabilism is unrivaled when it comes to this ameliorative function of epistemology (see our discussion of the planning fallacy, above, and chapters 3 and 9 of *EPHJ*). Some philosophers do not take this to be an advantage of SR. For these philosophers, aiming for a theory that provides useful guidance is ‘an objectionable project for epistemology’. Traditional epistemological theories ‘give no guidance at all’ because ‘[t]he advisability of holding a belief is independent of its justification’ (Conee 3; see also Feldman and Conee; Sosa, ‘Defense of Intuitions’). We’re inclined to think that these philosophers are right about theories of justification – they don’t provide useful epistemic guidance. What we don’t see is why this makes traditional theories preferable to SR. *Ceteris paribus*, a normative theory that evaluates and recommends beliefs is obviously superior to a theory that only evaluates beliefs. Perhaps SR makes poor epistemic evaluations or offers lousy epistemic recommendations. We expect objections to this effect. But we really don’t understand the objection that our theory is just too powerful – it does too much. What drives so many philosophers to loudly insist upon the practical irrelevance of epistemology? We see no good reason to legislate disciplinary boundaries so as to rule out a priori an ameliorative, recommendatory function for epistemology.<sup>2</sup>

Yet another objection along these general lines is that insofar as the evidence for SR comes from successful applications of various reasoning methods – or ‘epistemic casuistry’ –

there is no unified discipline there. Such casuistry would encompass all the manuals for all the various instruments and how to read all the various gauges, for one thing. And it would also include the variegated practical lore on how to tell what’s what and on what basis: the lore of navigation, jungle guidance, farming tips, and so on and so forth. That is all of course extremely useful, but it is no part of the traditional problematic of epistemology. (Sosa, ‘Experimental Philosophy’ 106)

We don’t know much about navigation, jungle guidance, or farming, so we can’t really say what sorts of lessons, epistemological or otherwise, one might learn from a close study of those practices. What we *are* committed to is just this: People reason in an epistemically excellent fashion about these (significant) matters to the extent that they reason in a robustly reliable and efficient manner. If SR plausibly accounts for the epistemic quality of the reasoning behind this ‘variegated practical lore’ – and we think it does – then there is no reason to question whether there is a ‘unified discipline’ here. Strategic Reliabilism is the theory that unifies this discipline.

#### 2.4. GOOD REASONING BY CHANCE

Strategic Reliabilism is bound to attract a certain class of objections that rely on improperly acquired but highly reliable cognitive processes

(BonJour, *Structure of Empirical Knowledge*; Plantinga). Would reasoning produced by a reliable cognitive mechanism that arose by luck (e.g., a brain lesion) count as high quality reasoning? What if a reasoner chooses a particular reasoning strategy as a result of poor reasoning (e.g., wishful thinking) but that strategy is, by pure happenstance, highly reliable? Recall that SR must provide an account of the epistemically proper acquisition and revision of cognitive mechanisms. According to SR, the cognitive mechanism that produces a belief is excellent only if that mechanism was properly acquired. A natural way to account for this would appeal to second-order reliability: There is some process that explains how we acquired the cognitive mechanisms that we have. A cognitive mechanism is properly acquired if and only if it is acquired as a result of a process-type that tends to produce reliable mechanisms. Broadly speaking, there are two ways to acquire or revise a cognitive mechanism.

1. By reflective reasoning. When reflective reasoning is poor (e.g., a reasoner adopts a new cognitive mechanism as the result of inappropriate guesswork), then the mechanism was not properly acquired, even if it is highly reliable. A mechanism that is acquired as the result of high quality reflective reasoning is properly acquired.
2. By processes outside our conscious control. Here too we want to distinguish between proper acquisitions (e.g., the result of the normal course of cognitive development) and improper ones (e.g., a brain lesion). So a mechanism is properly acquired if it is the result of the natural course of cognitive development; but it is improperly acquired if it is the result of (say) a brain lesion.

So a cognitive mechanism that produces a belief, considered in isolation, might be epistemically excellent. But it's still possible that the entirety of the reasoning that produced that belief was not epistemically excellent because that mechanism was not properly acquired. When an inappropriately acquired but reliable cognitive mechanism is used successfully, it is possible that over time, its status might change. But we will leave these complexities for another time.

## 2.5. THE ROLE OF INTUITIONS IN NATURALIZED EPISTEMOLOGY

In *EPHJ*, we criticized the method implicit in a lot of contemporary epistemology, which involves building theories that capture our 'intuitions' about knowledge, justification, and other philosophically significant categories. There has been considerable debate over the proper status of these intuitions. Many naturalistically inclined epistemologists have been quite pessimistic about intuitions, and we certainly share this pessimism. Some have argued that pessimism about our epistemic intuitions leads to epistemological apocalypse (see BonJour, *In Defense of Pure Reason* 99; Bealer). The gist of this objection is stated nicely by Mark Kaplan.

I am . . . struck by the fact...that what the proponents of epistemological naturalism have offered us is a series of arguments. And I am struck by the fact that they have not offered us any scientific evidence that, were we to believe what they would have us believe as a causal consequence of their having exposed us to these arguments, our belief would be the result of our having instantiated a process that is apt to have a high frequency of generating epistemically virtuous states in human beings in our world . . . It is hard to see what we can do except evaluate these arguments by the light of the very sorts of epistemic intuitions which the naturalists are so eager to disparage. (359–60)

We believe this argument (as well as its variations) attack a strawman. Most epistemological naturalists do not disparage *all* epistemic intuitions (e.g., Kornblith, *Knowledge and Its Place*). Epistemological naturalists tend to disparage epistemological *theories* and *traditions* that as a methodological matter take capturing our epistemic intuitions (e.g., about knowledge and justification) to be the primary – and perhaps only – substantive desideratum on a successful epistemological theory. From our perspective, this places more theoretical weight on intuitions than they can reasonably bear. Why? We think that a plausible and robust epistemological theory must be capable of offering effective recommendations for how people ought (epistemically) to reason and what people ought (epistemically) to believe. Some philosophers deny this, as we have noted, by insisting that epistemology be silent about these matters. But we refuse to make the prescriptive impotence of epistemology a pillar of our philosophical method. We refuse to comfort the philistines who snicker that philosophy has nothing to tell us about how to live by explicitly building that assumption right into our practice. So given our prescriptive aspirations, we have a very good reason not to rely exclusively on our intuitions in building our epistemological theories: As an empirical matter, our epistemic intuitions are not as reliable as science at identifying reliable belief-forming processes. That's not to say that our epistemic intuitions are totally unreliable, or that we may never legitimately appeal to our intuitions. In fact, SR together with empirical findings about how effectively our epistemic intuitions identify tractable, robustly reliable reasoning strategies might render useful general advice about when we ought (epistemically) to trust our intuitions about how we ought to reason and what we ought to believe.

An epistemological theory that renders judgments that give us the best chance to believe significant truths is preferable to a theory that renders judgments that respect our epistemic intuitions. If SR recommends a cognitive process whose employment would controvert some of our firmly held epistemic intuitions, we would be 'eager to disparage' these particular intuitions (Bishop, 'In Praise of Epistemic Irresponsibility'; Trout, 'Scientific Explanation'; 'Psychology of Scientific Explanation'). Our intuitions, philosophical or otherwise, deserve a quick death when they stand in the way of our reasoning in robustly reliable ways about significant problems. This methodological commitment does not lead to Epistemological

Apocalypse. We think it is in fact – dare we say it? – quite intuitive. More importantly, it is a commitment that is essential to what we believe are the legitimate prescriptive aspirations of epistemology.

## 2.6. WHY NOT PRAGMATISM?

In defending SR, we have appealed to the Aristotelian Principle: In the long run, good reasoning tends to lead to good outcomes. One might wonder why we don't take the Aristotelian Principle to its natural extreme and embrace pragmatism. Stephen Stich presses precisely this point:

[I]n some very significant situations, having false beliefs leads to better outcomes than having true beliefs. Though examples are legion, perhaps the best known comes from the work of Shelley Taylor and her colleagues who have shown that 'positive illusions' and 'unrealistic optimism' in patients with HIV leads to both better psychological coping and slower progression of the infection. To put the matter simply, if you have false beliefs you live longer and have a higher quality of life. Other investigators have found similar results in patients with heart disease. This suggests that in trying to extract insights about 'epistemic excellences' from Ameliorative Psychology in the Meehl and Gigerenzer traditions, [Bishop & Trout] have too narrow a focus. If they take the Aristotelian Principle seriously, then, at least in some domains, good reasoning will be robustly *unreliable*. ('Review' 392–3)

It would not be surprising, in view of the evolutionary gumbo that humans are, if having a reliable cognitive mechanism is sometimes maladaptive, or having an unreliable cognitive mechanism is sometimes conducive to well-being. But we think it is a mistake to suppose that the Aristotelian Principle (good reasoning leads to good outcomes) is a necessary or conceptual truth rather than a contingent one.

The view Stich is pressing embraces 'the very Jamesian contention that *there are no intrinsic epistemic virtues*' (*Fragmentation of Reason* 24). 'For pragmatists, there are *no* special cognitive or epistemological values. There are just *values*' ('Naturalizing Epistemology' 9). This is to be sharply distinguished from our view. Strategic Reliabilism makes distinctively epistemic recommendations – it tells us what are the *epistemically* best ways we can reason. On our view, there are 'intrinsic epistemic virtues'. It's true that SR explicitly employs some pragmatic factors: it recommends *tractable* reasoning strategies that tackle *significant* problems. But this is because an epistemological theory with prescriptive bite must recognize and compensate for our finite capacities. Strategic Reliabilism does not reduce to pragmatism. As Stich notes, false beliefs can sometimes yield a higher quality of life. In these cases, SR might yield *prima facie* recommendations about how to reason that compete with – and can be overridden by – other normative (e.g., moral or prudential) considerations.

The view Stich is pressing denies the existence of boundaries between what we normally think of as different areas of evaluation – e.g., epistemic,

prudential, moral, aesthetic. We resist this glomming together of values into one great evaluative package. There are three reasons (including two pragmatic ones) to keep the epistemic domain separate from other evaluative domains. First, failing to do so will result in an unwieldy epistemology. Strategic Reliabilism evaluates and recommends cognitive mechanisms primarily in terms of their robust reliability, while Stich's pragmatism evaluates and recommends cognitive mechanisms primarily in terms of their expected utility for the reasoner (where this is a function of the entirety of the reasoner's intrinsic values). But surely it's easier to generalize about reliable mechanisms than about useful ones. People's intrinsic values are so varied that we do not see how a theory can offer useful general guidance to reasoners of the sort SR can commonly deliver. For example, using a tractable Bayesian reasoning strategy to tackle diagnostic problems is the most reliable reasoning strategy available; SR tells us that this is the epistemically best way for us to reason about diagnostic problems (*EPHJ* 139–44). And so *ceteris paribus*, this is how we ought to reason. But does the Bayesian reasoning strategy always maximize the reasoner's expected utility? Probably not. So SR can offer more general, useful advice than pragmatism. What's more, the usable advice SR delivers tends to lead to positive outcomes. Thus, SR should be preferred to pragmatism on pragmatic grounds: It can offer more useful advice that leads to superior pragmatic outcomes.

The second reason we have for being reticent to embrace pragmatism depends on a pair of empirical hypotheses.

1. Many of our actual epistemic practices involve evaluating beliefs and reasoning in truth-linked terms (how likely is the belief to be true? how much evidence do we have for thinking the belief is true? how reliable is that reasoning?).
2. It would be very difficult for most people and institutions to overthrow those truth-linked practices for evaluating beliefs and reasoning and replace them with entirely pragmatic practices for evaluating beliefs and reasoning.

Suppose these empirical hypotheses are true. Suppose further that the correct epistemological theory should be able to be usefully applied by people and institutions. (This assumption will surely be granted by a pragmatist.) In that case, once again, we have a pragmatic reason to construct a theory for evaluating beliefs and reasoning that considers distinctly truth-linked, epistemological values (Bishop, 'Reflections on Cognitive and Epistemic Diversity').

We suspect that the above reasons we have for rejecting pragmatism – it is too unwieldy and too alien to be effectively implemented – derives from a deeper truth: the distinctions between evaluative domains descend from the kinds of capacities that we humans have, what the world is like, and the systematic relations between the world and our capacities. For example, the prudential domain is tied to considerations of human

well-being (as is the moral domain, at least in part). And we take well-being to be a complex property of humans that psychology measures and studies. The epistemic domain is tied to considerations of effectively gathering accurate and useful information about the world. And the moral domain is tied to considerations of social cooperation (among, perhaps, others). These evaluative domains reflect different aspects of our natures that are independent and not reducible one to the other. Nonetheless, as a contingent matter, we live in a world in which these considerations usually ‘point’ in the same direction – in most situations, the distinct demands of good reasoning, morality and prudence bind us in coherent and overlapping ways. When morality demands honesty, the good reasoner will tend to communicate more significant truths, and this virtue will tend to promote one’s self interest. Of course, this normative consilience can break down. Philosophers tend to focus on cases where these demands are at odds – when, for example, the demands of morality compete with those of prudence or of good reasoning. And while that focus is legitimate and understandable, it should not blind us to the overall normative coherence of everyday life. The various normative domains are distinct and irreducible but as a contingent matter tend to lead us in the same directions.

Our final reason for embracing a truth-based epistemological theory is that truth is likely to play a central explanatory role in accounting for our individual and social cognitive practices. What explains the utility of most beliefs is that they are true. In engineering, buildings stand better when your beliefs, and the reasoning strategies in which they are deployed, get the facts right about the strength of materials. In the history of science and technology – from rope and peg methods of construction to the domestication of plants and animals to finding cures for diseases – the usefulness of methods is explained in terms of the extent to which they produce accurate beliefs. Given the normative coherence of everyday life, it is not surprising that efforts to discern the truth about ourselves and the world is a remarkably good way to arrive at information we’ll find useful.

### 3. *Conclusion*

Strategic Reliabilism is a theory that is able to evaluate the epistemic quality of cognitive mechanisms and prescribe (from an epistemic perspective) various ways of forming beliefs. The core of the theory says that epistemically excellent reasoning is efficient reasoning that leads in a robustly reliable fashion to significant, true beliefs. The theory also provides an account of what it is for a cognitive mechanism to be legitimately acquired. And the theory will have an applied component that recommends various ways people can improve their reasoning. Some have worried that this applied component is but a dream – a set of principles that will be available far in some future brave new world. But for those who can be satisfied with something less than a pristine final theory, SR together with

the findings of psychology has more to offer than one might suppose. We can usefully envision the applied component of this theory as rendering two sorts of judgments.

1. Its conservative judgments identify a reasoner's causally operative reasoning mechanisms that can't be changed or are so epistemically excellent (compared to alternatives) that they should not be changed.
2. Its revisionary judgments identify replacement cognitive mechanisms and strategies – new ways of reasoning that the reasoner should be using to think about the world.

A lot of our cognitive machinery is hard to change, and so this part of the applied component will be – and has been – fruitfully investigated and described by psychology. As for the revisionary judgments, we still have a lot to learn. But we already know quite a bit. In *EPHJ* and in this paper, we have suggested many practical ways to improve how we reason about some pretty significant matters. Philosophers might sniff about how this advice is too concrete to yield useful theoretical insights or how it won't help us resolve deep scientific or philosophical conundrums. But the planning fallacy doesn't just lead us to wildly overbudget opera houses. In March 2003, polls in the United States showed 70% support for the invasion of Iraq – a war that was billed a 'cakewalk' by one of its prominent supporters, a war that members of the Bush administration predicted would last 'weeks rather than months' and would cost 'under \$50 billion'.<sup>3</sup> Five years (and counting) later, we want more than a theory that can at best note that these tragically optimistic forecasts were epistemically unjustified. Surely there is a place in epistemology for a normative theory that can condemn our defective habits of mind and recommend strategies for improving upon them. Surely there is a place for a theory like Strategic Reliabilism.

### *Short Biographies*

Michael A. Bishop is Professor of Philosophy at Florida State University. He has authored or co-authored articles on a wide range of issues in philosophy of science and epistemology, which have appeared in venues such as *Philosophy of Science*, *Nous*, *Philosophical Studies*, and *Synthese*. In philosophy of science, Bishop has addressed the nature and role of thought experiments in science, theory-ladenness of perception, semantic incommensurability and the realism–antirealism debate. In epistemology, he has addressed issues concerning the nature of epistemic responsibility, naturalism in epistemology, the 'rationality wars', and social epistemology. He is co-author, with J. D. Trout, of *Epistemology and the Psychology of Human Judgment* (Oxford, 2005). Bishop has held a National Endowment for the Humanities Summer Fellowship and a National Science Foundation Scholars Award. Before arriving at Florida State University, Bishop taught

at Iowa State University and Northern Illinois University. He holds a B.A. in Philosophy from the University of Maryland, and a Ph.D. in Philosophy from the University of California, San Diego.

J. D. Trout works at the borders of philosophy of science, epistemology, and psychology. He has published in such journals as *Philosophy of Science*, *Nous*, *Psychological Review*, *Trends in Cognitive Sciences*, *Current Directions in Psychological Science*, and *British Journal for the Philosophy of Science*. His experimental research in speech perception has been published in *Language and Speech* and *Speech Communication*. He has authored or co-authored a number of books. *Measuring the Intentional World* (Oxford, 1998) argues that progress in the social and behavioral sciences warrants a measured argument for scientific realism in those domains. *Epistemology and the Psychology of Human Judgment* (with Michael Bishop; Oxford, 2005), proposes that the Standard Analytic Epistemology practiced in the English-speaking world should be replaced by Ameliorative Psychology. Trout's book, *The Empathy Gap* (now in press, Viking/Penguin, 2009), highlights the promise of judgment research for the improvement of human well-being. He has held a National Science Foundation pre-doctoral fellowship, a Mellon post-doctoral fellowship, a National Endowment for the Humanities Summer Fellowship, and a National Science Foundation Scholars Award. Before arriving at Loyola University Chicago, Trout taught at Bryn Mawr College and Virginia Tech, and he has visited at the University of Innsbruck and the University of Chicago. He holds a B.A. in Philosophy and History from Bucknell University, and a Ph.D. in Philosophy from Cornell University.

### Notes

\* Correspondence address: J. D. Trout, Loyola University Chicago, 6525 N. Sheridan Road, Chicago, IL 60626, USA. Email: jtrout@luc.edu.

<sup>1</sup> The reviewer raised this objection.

<sup>2</sup> A historical case for this restriction won't move us, since we are not inclined to expunge figures such as Descartes (in *Rules for the Direction of the Mind*), Bacon, Mill and Bayesians from the ranks of legitimate epistemologists.

<sup>3</sup> According to Gallup polls, 73% believed US action in Iraq was 'morally justified' (March 29–30) and 72% favored the war in Iraq (March 22–23) <<http://www.gallup.com>>. Ken Adelman billed the war as a 'cakewalk' twice in the *Washington Post* – first in 'Cakewalk in Iraq', Wednesday, February 13, 2002, p. A27 and again in '“Cakewalk” Revisited', Thursday, April 10, 2003, p. A29 <<http://www.washingtonpost.com/ac2/wp-dyn/A1512-2003Apr9>>. Vice President Cheney predicted the war would last 'weeks rather than months' on CBS's 'Face the Nation' on March 16, 2003 <<http://www.cbsnews.com/stories/2003/03/17/ftn/main544228.shtml>>. Secretary of Defense Donald Rumsfeld, estimated that a war in Iraq would cost 'something under \$50 billion' at a 'media stakeout' on January 19, 2003 <<http://www.defenselink.mil/transcripts/transcript.aspx?transcriptid=1322>>.

### Works Cited

Bealer, G. 'The Incoherence of Empiricism'. *Naturalism: A Critical Appraisal*. Eds. S. Wagner and R. Warner. Notre Dame UP, 1993. 163–96.

- Bishop, M. 'Fast and Frugal Heuristics'. *Philosophy Compass* 1.2 (2006): 201–23.
- . 'In Praise of Epistemic Irresponsibility: How Lazy and Ignorant Can You Be?' *Synthese* 122 (2000): 179–208.
- and J. D. Trout. *Epistemology and the Psychology of Human Judgment*. New York, NY: Oxford UP, 2005.
- . 'The Pathologies of Standard Analytic Epistemology'. *Notis* 39.4 (2005): 696–714.
- Bishop, Michael. 'Reflections on Cognitive and Epistemic Diversity: Does a Stich in Time Save Quine?'. *Stephen Stich and His Critics*. Eds D. Murphy and M. Bishop. Blackwell Press, forthcoming.
- BonJour, Laurence. *In Defense of Pure Reason*. Cambridge, UK: Cambridge UP, 1998.
- . *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard UP, 1985.
- Conee, E. 'Review of Epistemology and the Psychology of Human Judgment'. *Philosophy and Phenomenological Research*. Forthcoming. [Page numbers cited are from unpublished manuscript.]
- and R. Feldman. 'The Generality Problem for Reliabilism'. *Philosophical Studies* 89 (1998): 1–29.
- de Neufville, Richard. 'Airport Privatization: Issues for the United States'. *Safety, Economic, Environmental, and Technical Issues in Air Transportation*. Transportation Research Record 1662, Paper 99.0218, 1999. 24–31.
- . 'The Baggage System at Denver: Prospects and Lessons'. *Journal of Air Transport Management* 1.4 (1994): 239–46.
- Dijksterhuis, Ap, M. W. Bos, L. F. Nordfren, and Rick B. van Baaren. 'On Making the Right Choice: The Deliberation-Without-Attention Effect'. *Science* 311 (2006): 1005–7.
- Eikeland, Per T. 'The Preconditions of Cost Control in Building Projects'. <<http://www.pte.no/assets/documents/publikasjoner/preconcost.pdf>>, accessed 2 July 2008.
- Feldman, R. 'Reliability and Justification'. *The Monist* 68 (1985): 159–74.
- and E. Conee. 'Evidentialism'. *Philosophical Studies* 48 (1985): 15–34.
- Flyvbjerg, Bent. 'Machiavelian Megaprojects'. *Antipode: A Radical Journal of Geography* 37 (2005): 18–22.
- Gigerenzer, G. *Gut Feelings*. New York, NY: Penguin, 2007.
- Gladwell, M. *Blink*. New York, NY: Little Brown, 2005.
- Goldman, Alan. 'Review: Epistemology and the Psychology of Human Judgment'. *Notre Dame Philosophical Reviews*. 2005. <<http://ndpr.nd.edu/review.cfm?id=2701>>.
- . 'What is Justified Belief?' *Justification and Knowledge*. Ed. G. Pappas. Dordrecht: D. Reidel, 1979. 1–23.
- Goldman, Alvin I. 'Internalism Exposed'. *Journal of Philosophy* 96 (1999): 271–93.
- Hogarth, R. and P. Schoemaker. 'Beyond Blink: A Challenge to Behavioral Decision-Making'. *Journal of Behavioral Decision-Making* 18 (2005): 305–9.
- Kahneman, D. and A. Tversky. 'Intuitive Prediction: Biases and Corrective Procedures'. *Judgment under Uncertainty: Heuristics and Biases*. Eds. D. Kahneman, P. Slovic, and A. Tversky. Cambridge, UK: Cambridge UP, 1979/1982. 414–21.
- Kaplan, M. 'Epistemology Denatured'. *Midwest Studies in Philosophy* 19 (1994): 350–65.
- Kornblith, H. 'Justified Belief and Epistemically Responsible Action'. *The Philosophical Review* 92 (1983): 33–48.
- . *Knowledge and Its Place in Nature*. Cambridge, MA: MIT Press, 2002.
- Morris, M. and R. Nisbett. 'Tools of the Trade: Deductive Schemas taught in Psychology and Philosophy'. *Rules for Reasoning*. Ed. Richard Nisbett. Hillsdale, NJ: Erlbaum, 1993. 228–56.
- Neta, Ram. 'How to Naturalize Epistemology'. *New Waves in Epistemology*. Eds. V. Hendricks and D. Pritchard. New York, NY: Palgrave, 2007. 324–53.
- Pollock, John and Joseph Cruz. *Contemporary Theories of Knowledge*. Lanham, MD: Rowman & Littlefield, 1999.
- Plantinga, A. *Warrant: The Current Debate*. New York: Oxford University Press, 1993.
- Redelmeier, D. and D. Kahneman. 'Patient's Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluation of Two Minimally Invasive Procedures'. *Pain* 66 (1996): 3–8.
- Scott, James, Margaret Stumpp, and Peter Xu. 'Overconfidence Bias in International Stock Prices: Consistent Across Countries and Trading Environments'. *The Journal of Portfolio Management* (Winter 2003): 80–9.
- Sosa, Ernest. 'A Defense of Intuitions'. *Stephen Stich and His Critics*. Eds. Dominic Murphy and Michael Bishop. Blackwell, forthcoming.

- . ‘Experimental Philosophy and Philosophical Intuitions’. *Philosophical Studies* 132 (2007): 99–107.
- Stich, Stephen P. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: Bradford Books/MIT Press, 1990.
- . ‘Naturalizing Epistemology: Quine, Simon and the Prospects for Pragmatism’. *Philosophy and Cognitive Science*. Royal Institute of Philosophy, Supplement no. 34. Eds. C. Hookway and D. Peterson. Cambridge: Cambridge UP, 1993. 1–17.
- . ‘Review: Epistemology and the Psychology of Human Judgement’. *Mind* 115 (2006): 390–3.
- Trout, J. D. *The Empathy Gap*. New York, NY: Viking/Penguin, forthcoming 2009.
- . ‘The Psychology of Scientific Explanation’. *Philosophy Compass* 2.3 (2007): 564–91.
- . ‘Scientific Explanation and the Sense of Understanding’. *Philosophy of Science* 69.2 (2002): 212–33.
- Tykocinski, O. ‘I Never Had a Chance: Using Hindsight Tactics to Mitigate Disappointments’. *Personality and Social Psychology Bulletin* 27.3 (2001): 376–82.
- Van Fraassen, B. ‘The Pragmatics of Explanation’. *American Philosophical Quarterly* 14 (1977): 143–150.
- Weinberg, J., S. Nichols, and S. Stich. Normativity and Epistemic Intuitions. *Philosophical Topics* 29.1–2 (2001): 429–60.