Advanced Review

Rationality

Clifford Sosis and Michael A Bishop*

A theory of rationality is a theory that evaluates instances of reasoning as rational, irrational, or (ir)rational to some degree. Theories can be categorized as rule-based or consequentialist. Rule-based theories say that rational reasoning accords with certain rules (e.g., of logic or probability). Consequentialist theories say that rational reasoning tends to produce good consequences. For instance, the reliabilist takes rationality to be reasoning that tends to produce mostly true beliefs. The pragmatist takes it to be reasoning that tends to produce mostly useful beliefs. This article reviews some of the features and the challenges of rule-based, reliabilist, and pragmatist theories of rationality.

How to cite this article:
WIREs Cogn Sci 2013. doi: 10.1002/wcs.1263

INTRODUCTION

Rationality is a property of reasoning. An instance of reasoning might be rational, irrational, or it might be rational to some degree. Philosophers typically distinguish between theoretical rationality and practical rationality. Theoretical rationality applies to reasoning that is capable of producing belief (e.g., that Socrates is a man). Practical reasoning applies to reasoning that is capable of producing action or the intention to act (e.g., purchase the video game for $49). Here we will be concerned with theoretical rationality.

EMPIRICAL STARTING POINTS

Many contemporary discussions of rationality can be understood to be reactions to four different lines of empirical research.

Heuristics and Biases

The basic lesson of the Heuristics and Biases (HB) program is that (a) we are naturally disposed to use simple reasoning strategies, and (b) these strategies are often less reliable in the long run than ideal rules. Some have argued that this line of research has 'bleak implications' for human rationality.[1] Many HB program examples are so familiar that the cognoscenti refer to them with nicknames—e.g., the Linda Problem (aka the Conjunction Fallacy), base rate neglect. Let's begin with the Linda problem.[2]

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.
>
> Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.
>
> (a) Linda is a teacher in elementary school.
>
> (b) Linda works in a bookstore and takes Yoga classes.
>
> (c) Linda is active in the feminist movement.
>
> (d) Linda is a psychiatric social worker.
>
> (e) Linda is a member of the League of Women Voters.
>
> (f) Linda is a bank teller.
>
> (g) Linda is an insurance sales person.
>
> (h) Linda is a bank teller and is active in the feminist movement.

Most people judge (h) to be more likely than (f) (Ref 2, p. 93). This is so even when they are given only two options (f, h) (Ref 2, p. 94). This judgment is taken to be problematic because it violates the laws

*Correspondence to: mbishop@fsu.edu

Department of Philosophy, Florida State University, Tallahassee, FL, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.

of probability. The probability that a conjunction is true (e.g., Barack Obama is wearing socks and Barack Obama is wearing shoes) can never be greater than the probability that either conjunct alone is true. And so the probability that Linda is both a bank teller *and* a feminist cannot be greater than the probability that she is a bank teller.

The second example we will consider is base rate neglect. The following scenario was presented to 60 students and staff at Harvard Medical School (Ref 3, p. 999):

> If a test to detect a disease whose prevalence is 1/1,000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs?

This description omits a piece of information essential to solving the problem: the sensitivity of the test (i.e., the true positive rate). Assuming a true positive rate of 100%, the probability that the person has the disease is about 2%. But few people come to this conclusion. Many simply use the test's accuracy to estimate the probability that someone who tests positive has the disease. So almost half of the students and staff at Harvard Medical School judged the probability to be 95%; the mean answer was 56%; and only 18% gave the answer sanctioned by the rules of probability. (Although the above study omits the test's sensitivity, other studies that include this information have come to very similar results.[4,5])

Predictive Modeling

The basic lesson of this literature is that (a′) there are simple reasoning strategies we are not naturally disposed to use, and (b′) these strategies are about as reliable as ideal ones. In fact, these models are more reliable than experts who have access to the same evidence.[6] Examples include models for predicting marital happiness,[7] criminal recidivism,[8] academic performance,[9–11] credit risk,[12] Sudden Infant Death Syndrome,[13,14] and the quality of red Bordeaux wines.[15]

Fast and Frugal Heuristics

This line of research can be understood as a reaction to the above two literatures. It embraces half of each of their pessimistic lessons, for a rather optimistic result: (a) we are naturally disposed to use simple reasoning strategies, and (b′) these strategies are about as reliable as ideal ones.[16] The Fast and Frugal Heuristics program (FFH) prescribes a whole host of reasoning strategies that are fast (i.e., easy to use and allow us to make judgments quickly) and frugal (i.e., rely on a fraction of the available evidence). Some of these heuristics are controversial. But a very nice FFH result is that people can improve their reasoning on diagnosis problems by framing them in terms of frequencies rather than probabilities. So consider the Harvard Medical School problem. Rather than frame the problem in terms of false positive rate (e.g., 5%), the trick is to suppose you give the test to 1,000 people, one of whom has the disease. The person with the disease will test positive. And about 50 people without the disease (5% of 999) will also test positive. And so of the (roughly) 51 people who test positive, only one will actually have the disease. What Gigerenzer and Hoffrage found was that when diagnosis problems were framed in terms of frequencies rather than probabilities, most people reasoned in accordance with Bayes' rule.[5]

Diversity

There appear to be significant and systematic differences in how different people reason about the world. And '[i]t is the prospect of cognitive diversity among normal folk that lends a genuine, almost existential urgency to the project of cognitive evaluation' (Ref 17, p. 74). This is implicit in the above studies—different people reason in different ways about the Linda problem.[18] In a series of studies, Richard Nisbett and his colleagues have argued that 'literally different cognitive processes are often invoked by East Asians and Westerners dealing with the same problem' (Ref 19, p. 305). These differences are explained in terms of deep, long-standing cultural differences between East and West.[19] Indeed, 'Asians move radically in an American direction after a generation or less in the United States' (Ref 19, p. 307). To take just one example, Peng and Nisbett showed that in the face of conflicting claims, Westerners tend to become more confident in the statement they deem more plausible, whereas East Asians become less confident.[19] East Asians insist on a 'Middle Way' while Westerners insist on 'My Way'.[20] (For another fascinating and important study on cognitive diversity, see Ref 21.) In the face of the Linda problem or the base rate neglect problem, different people reason in different ways. Is someone reasoning rationally? Irrationally? Why should we think so? Westerners and Asians reason in very different ways about a whole host of problems. Whose reasoning is better? And why? To answer these questions, we need some idea of what it is to reason rationally and irrationally. Our plan is to introduce three families of theories of rationality. Two are consequentialist.

They evaluate cognitive mechanisms in terms of their consequences. The third family consists of rule-based theories. They evaluate cognitive mechanisms in terms of how well they accord with certain reasoning rules. We will begin our review with a rule-based view.

THE STANDARD PICTURE OF RATIONALITY

A natural explanation for why someone's reasoning might be irrational is that it violates certain well-established inferential rules. Edward Stein has called this the 'Standard Picture' of rationality:

> According to this picture, to be rational is to reason in accordance with principles of reasoning that are based on rules of logic, probability theory and so forth. If the standard picture of reasoning is right, principles of reasoning that are based on such rules are normative principles of reasoning, namely they are the principles we ought to reason in accordance with (Ref 22, p. 4).

In the Linda problem, people reason irrationally when they violate the laws of probability (by judging the probability of A and B to be greater than the probability of A). In base rate neglect problems, people reason irrationally when they violate Bayes' rule (which can be derived from standard axioms of probability theory).

Regardless of whether psychologists would, upon reflection, embrace the Standard Picture, they sometimes appear to presuppose it in explaining why people are reasoning irrationally. For example, Kahneman and Tversky explain what makes something an error of judgment: 'The presence of an error of judgment is demonstrated by comparing people's responses either with an established fact . . . or with an accepted rule of arithmetic, logic, or statistics' (Ref 2, p. 124). Massimo Piattelli-Palmarini argues that people make the above sorts of errors because they suffer from 'probability blindness' (Ref 23, p. 130–132). And the article by Gigerenzer and Hoffrage that shows how to improve doctors' reasoning about diagnosis problem is called 'How to *Improve* Bayesian Reasoning Without Instruction: Frequency Formats' (emphasis added).[5] Implicit in the title is the assumption that people's reasoning improves when it is in accordance with Bayes' rule.

We should note that just because one explains why some piece of reasoning is irrational by noting that a rule of logic or probability has been violated, she is not thereby committed to the Standard Picture.[6,24] One of us (MB) is a reliabilist (see *Reliabilism* section). He thinks (roughly) that rational reasoning involves the operation of reliable (i.e., truth-conducive) cognitive mechanisms. If a cognitive mechanism systematically violates rules of logic or probability, that might be excellent evidence that the mechanism is unreliable (it does not produce mostly truths). One reason the reliabilist might reject the Standard Picture is that rule violations and reliable reasoning sometimes part ways. Consider that even the most successful scientific theories face anomalies, i.e., observations that violate the theory's predictions. The reliabilist might endorse both the theory and the observations, if both are the products of reliable belief-forming mechanisms, even though their conjunction implies an inconsistency. Scientific practice seems to side with the reliabilist on at least some of these episodes. (For an example, see *Problems with Bayesianism and the Standard Picture of Rationality* section.)

CONCEPTUAL REJECT-THE-NORM ARGUMENTS

Some have responded to allegations of widespread human irrationality with a conceptual reject-the-norm argument (RNA), which aims to show that the very nature of rationality makes it impossible for most people to be irrational (Ref 22, p. 239–242). As a result, if the Standard Picture (or any other theory of rationality) implies that most people are irrational, it must be employing the wrong norms. Most people's reasoning on the Linda problem or on the base rate problem cannot be irrational. Proponents of RNA admit that people, even most people, might reason poorly on certain problems. So it might be the case that most people's actual reasoning on (say) the base rate problem is poor and their judgments wrong. But that is not because people are irrational. They do not lack the *competence* to reason rationally about the problem. Rather, their errors are *performance* errors—the result of a failure of attention, memory, brute cognitive power, or comprehension (i.e., not properly understanding the problem).

There are various versions of RNA. L. J. Cohen put forward a RNA over three decades ago that has received considerable attention. Cohen argued that 'ordinary human reasoning—by which I mean the reasoning of adults who have not been systematically educated in any branch of logic or probability theory—cannot be held to be faultily programmed: it sets its own standards' (Ref 25, p. 317). Gerd Gigerenzer has defended a line of argument that has often been interpreted as a RNA.[26] He contends that people's reasoning in the Linda case and other similar cases are not errors because they do not violate the axioms of probability given

a frequentist interpretation of probability. According to that interpretation, single-event probabilities are meaningless. In the Linda problem, people are asked to assess the probability of single events (e.g., whether Linda is a bank teller). Gigerenzer argues: 'what is called the 'conjunction fallacy' is a violation of some subjective theories of probability, including Bayesian theory. It is not, however, a violation of the major view of probability, the frequentist conception' (Ref 26, p. 92). Gigerenzer's argument has produced a number of interesting exchanges[27–29] and some pointed criticism (Ref 24, p. 123–127). One of us (MB) has argued that Gigerenzer's argument is not best understood as a defense of the Standard Picture, but as a *reduction ad absurdum* of it (Ref 24, p. 127–133).

We propose to focus on a less picked over RNA defended by Sosa and Galloway (S&G).[30] S&G argue that like height or visual acuity, rationality is a concept that is 'implicitly indexical' (Ref 30, p. 173). It is tied (or indexed) to our actual situation and abilities: 'when we assess rationality we do so relative to our actual levels of rationality' (Ref 30, p. 174). And therefore, 'it is conceptually impossible to demonstrate that humans generally are irrational. The most that might be shown is that humans are not as rational as they might be, which seems a much less dramatic result' (Ref 30, p. 174).

Compared to a super-reasoner, we might score rather low on the rationality scale. Just as with respect to visual acuity, 'humans are generally deficient relative to eagles, and relative to Superman' (Ref 30, p. 174). But given the nature of rationality, and the fact that its standards are indexed to normal human abilities, the fact that we are deficient relative to non-humans with respect to visual acuity, height, or rationality is irrelevant. To say that humans are generally and seriously irrational is:

> as conceptually confused as it would be to say that humans are generally, in actuality, seriously short. Not that it is confused to think that people might have been seriously short, all of them. The confusion resides rather in the thought that for all we know humans are *in fact*, *in actuality*, seriously short, all of them. (Ref 30, p. 174–175)

The idea here is that it is impossible for all people to be short (given our indexed concept of height) or irrational (given our indexed concept of rationality). But it is possible that all people might have been short (given our actual concept of height) or irrational (given our actual concept of rationality). Of course, if these potential people have a concept of height or rationality, their concept would be indexed to *their* situation and so given *their* concept of height or rationality they could not all be short or irrational. Once we understand that rationality is indexed to our actual rational abilities, it is literally incoherent to think that all or even most of us are actually irrational. The best that empirical work on our reasoning abilities could show is that our rational accomplishments are more modest than we might have believed or hoped on the basis casual experience. But no empirical evidence can possibly show that people are generally irrational.

This sort of Protagorean view of rationality (where 'man is the measure of all things') faces a pair of related problems. The first arises from the fact that there is diversity in how people reason. Recall that in the Linda and base rate neglect problems, some people reasoned in accordance with the Standard Model. Most did not. So how can we index rationality to 'normal human abilities' if there is no single 'normal' human ability? What if our rational abilities have a bimodal distribution? Indeed, Stanovich discovered that people who reason well (according to the Standard Model) on some problems also tend to reason well on other problems and score higher on various intelligence tests. There is a more dramatic possibility raised by the diversity literature. What if our rational abilities are multimodally distributed? If this is so, then any theory of rationality that assumes that human reasoning abilities are distributed so that we can easily identify a single set range of 'normal human [reasoning] abilities' is doomed on empirical grounds.

S&G assume that rationality is best understood as a natural human characteristic (akin to height or visual acuity) that is relatively fixed, absent medical or technological enhancements, and distributed approximately normally (i.e., a rough bell curve distribution). We have argued that as an empirical matter, our reasoning abilities might not be distributed in so simple a manner. In fact, our ability to reason might be more like a skill that can be developed and improved than a (relatively) fixed characteristic. It might be more like playing the saxophone than visual acuity. Some people can make beautiful music with a saxophone. But most people who put lip to sax would make a flatulent, offensive noise. It would be a mistake to suppose that the dissonant wind we make with a sax shows us to be good players on the grounds that we're no worse than most. There is at least some evidence that reasoning, like playing a sport or a musical instrument, is a skill that can be improved with proper training and practice. And some studies suggest that dramatic improvements can come with surprisingly little training.[5,31] The fact that philosophy departments teach critical thinking courses is implicit

acknowledgement that reasoning is a skill that can be improved with instruction. And if this is the right way to understand reasoning and the evaluation of reasoning, then the Protagorean standard imposed by reject-the-norm arguments is inappropriate.

BAYESIANISM

There are a number of quite different views that fly under the flag of Bayesianism.[32] Our plan is to take a simple, bare bones version of Bayesianism to be an exemplar of the Standard Picture of Rationality. We will then review a number of worries that have been raised about Bayesianism and the Standard Picture. It is important to keep in mind that these are objections to Bayesianism *as a theory of individual rationality*, of how individuals ought to reason. They are not meant to undermine Bayesianism understood as a theory of probability or as a theory of how some (or perhaps all) scientific hypotheses are (or should be) confirmed.

A basic, bare-bones Bayesian theory of rationality consists of three theses. The first is that beliefs come in varying degrees that can be represented with probabilities. For example, I am more confident that the sun will rise tomorrow than I am that the Nationals will win the upcoming World Series. The second thesis is the coherence requirement. It says that a person's beliefs must satisfy the laws of probability. And the third thesis is the conditionalization requirement. It says that degrees of belief must be updated according to Bayes' rule. In its simplest form, the conditionalization requirement works as follows. Suppose at time t1 you believe A with probability D.

$$\Pr_1(A) = D$$

Suppose also that while you do not believe B, you believe A given B (A/B) with probability D' (where D' might be equal to D).

$$\Pr_1(A/B) = D'$$

Now suppose that at time t2 you come to believe B with certainty. The conditionalization requirement says that the new probability of A (i.e., the probability of A at time t2) is:

$$\Pr_2(A) = \Pr_1(A/B) = D'$$

The basic idea is that the probability you should assign to A after learning B is the probability you assigned A/B before learning B. This is a very simple form of Bayesianism. A more sophisticated version, for example, would allow for the possibility that you come to believe new evidence with something less than certainty.[33]

PROBLEMS WITH BAYESIANISM AND THE STANDARD PICTURE OF RATIONALITY

There are three problems with the Standard Picture of Rationality that also play out within a Bayesian framework. (1) Bayesian reasoning is not sufficient for rationality: one can reason in the Bayesian manner without reasoning rationally. (2) Bayesian reasoning is not necessary for rationality: one can reason rationally without reasoning in a Bayesian manner. And (3) the Bayesian requirements for rationality are too demanding: they require more cognitive resources than individual reasoners have.

Bayesianism requires that one's beliefs be coherent. But it seems possible to reason coherently but irrationally. For example, one might always assign probabilities to beliefs in a manner that is coherent with assigning a high probability to the belief that one is Napoleon. Consider that conditionalization works only after one has assigned the new evidence (B) a certain probability. But rational reasoning often involves assessing the quality of the new evidence (or of old background beliefs). Gilbert Harman notes that if one believes both P and *if* P *then* Q, this might give one good reason to believe Q. But it might also give one reason to give up P or *if* P *then* Q or both (Ref 34, p. 11–12). Harman asks us to suppose that Mary believes that if she looks in the closet she will see a box of Cheerios, she believes that she is looking in the closet, but she does not see a box of Cheerios. In this case, coherence requires only that we give up one of the beliefs. And in the normal case, it is not the observational belief (that she does not see a box of Cheerios) that should go but the conditional belief. There seem to be requirements on rationality—such as giving at least some prima facie weight to experience—that Bayesianism lacks. And so it is too weak.

Are the Bayesian requirements on reasoning necessary for rationality? For the Bayesian, it cannot be rational to believe (or to assign probability greater than 0.5 to) both P and not-P. But even our best scientific theories are sometimes inconsistent with our best observations. For example, in the 19th century, Newton's theory together with what was known about the solar system implied falsehoods about Mercury's orbit. And yet, it seems perfectly rational for (say) Maxwell to believe with a high degree of confidence (a) the laws of Newton's theory, (b) the accuracy of the observations of Mercury's orbit, and (c) that no object was exerting a force on Mercury that could account for the discrepancy between Newton's theory and Mercury's orbit. Newtonians knew perfectly well that (a) and (c) implied that (b) was false. And yet they

believed all three with a high degree of confidence. The Bayesian counts this as irrational; and intuitively, this seems wrong. (See also the discussion at the end of *The Standard Picture of Rationality* section.)

The third problem for Bayesianism as a theory of individual rationality is that it is psychologically unrealistic: It is practically impossible for a person to calibrate her degrees of belief to the axioms of probability. Consider that the laws of probability require that one not believe inconsistencies. Christopher Cherniak considers the resources it would take to check one's belief system for consistency via the truth table method.

> Suppose that each line of the truth table for the conjunction of all these beliefs could be checked in the time a light ray takes to traverse the diameter of a proton... and suppose that the computer was permitted to run for twenty billion years, the estimated time from the 'big-bang' dawn of the universe to the present. A belief system containing only 138 logically independent propositions would overwhelm the time resources of this supermachine (Ref 35, p. 93).

To see how this plays out for a Bayesian, suppose you believe that the Nationals will win the Series with a confidence level of 0.4, and that the Nationals will win the Series and Stephen Strasburg will win the Cy Young Award with a confidence level of 0.2. This commits you to assigning a specific degree of confidence (in fact, 0.5) to the belief that Strasburg will win the Cy Young Award given that the Nationals win the Series. But assigning probabilities in this way to every formally similar case would require more cognitive resources than we possess. The idea that an epistemological theory should respect our cognitive limitations is familiar in psychology.[36,37] It is more controversial among philosophers.[38,39] One motivation for adopting a consequentialist theory of rationality (such as reliabilism or pragmatism) is that they have the potential to provide guidance that does not require that we have brains 'the size of a blimp' (Ref 17, p. 27).

Nothing we have said here is meant to be a general critique of Bayesian reasoning. In particular contexts, Bayesian methods have proven to be extremely useful. Indeed, one of us (MB) is on record defending a Bayesian approach to reasoning about medical diagnoses.[24] Bayes' rule is an excellent tool for a rational person to have in her toolbox—as are various logical systems, arithmetic, geometry, and differential calculus. But if the worries raised against the Standard Picture are on track, these cannot be the only tools in the rational person's toolbox.

RELIABILISM

Among philosophers, reliabilism is taken to be a theory of justification—a theory about what makes an individual belief justified.[39] But insofar as reliability is a property of cognitive mechanisms, reliabilism is perhaps more naturally understood as a theory of rationality. Let's focus on a specific reliabilist theory of rationality, Bishop and Trout's (B&T) Strategic Reliabilism.[24]

B&T adopt a quirky approach to the study of rationality. Philosophers typically build theories that capture their understanding of an evaluative notion (e.g., justification, knowledge, rationality). The idea here is either that the goal of philosophy is to capture this understanding or that the goal of philosophy is to describe the real nature of (say) rationality and our understanding of rationality is highly faithful to its real nature. Following Stich,[17] B&T argue that neither of these approaches are likely to give us a motivated theory for evaluating how people ought to reason. And so, like Hilary Kornblith, who begins his epistemological theorizing with science, in particular, cognitive ethology,[40] B&T begin their epistemological theorizing with the empirical findings we canvassed in *Empirical Starting Points* section. They note that some of these lines of research are normative: they give explicit advice about how we ought to reason. B&T dub this research 'Ameliorative Psychology'. Rather than try to build a theory of rationality that answers to our (potentially flawed) commonsense understanding of rationality, B&T argue that we should embrace the normative framework that grounds the prescriptions of Ameliorative Psychology.

B&T argue that the normative judgments of Ameliorative Psychology spring from a reliabilist normative framework. This framework, Strategic Reliabilism (SR), holds that a cognitive process is rational to the extent it is (a) robustly reliable (i.e., reliable across a wide range of environments), (b) appropriate to the reasoner's resources, and (c) geared toward producing beliefs about topics that are significant for the reasoner. The best way to appreciate the power of SR is to see how it grounds the prescriptions of various parts of Ameliorative Psychology. Let's start with heuristics and biases. The simplest explanation for why it is irrational to (say) neglect base rates in one's diagnostic reasoning is that it systematically leads one to adopt inaccurate beliefs about very important matters (e.g., how many people in a population are HIV positive). In the predictive modeling literature, what recommends a simple model for predicting (say) criminal recidivism is that it is reliable—about as reliable as ideal (regression) models—and easy to use. It also focuses on a prediction problem ('Will this prisoner commit

another violent crime if paroled?') that is highly significant. The FFH program proposes heuristics that are purportedly robustly reliable, easy to use, and focused on significant problems. And these heuristics tend to be criticized on the grounds that they are unreliable or reliable only in very specific circumstances (i.e., not robust) or that they are relatively difficult to use. B&T contend that these normative considerations are grounded in a normative framework; and this normative framework is Strategic Reliabilism.

There are a number of tricky issues any reliabilist must face: What is a reasoning strategy? What is it for a reasoning strategy to be reliable? How reliable does a reasoning strategy have to be to be rational? The proponent of Strategic Reliabilism faces further challenges: What is it for a reasoning problem to be significant? What is it for a reasoning strategy to be easy or hard to use? And how do we trade off reliability against tractability (ease of use)? In other words, if one strategy is slightly more reliable but also more difficult to use than another, which is more rational? This is not the place to try to address these challenges. But it is worth noting that the reliabilist need not adopt a single answer to all these questions. Consider the cutoff problem: At what level of reliability is a reasoning strategy rational? We suggest that it is a mistake for the reliabilist to give a single answer (e.g., 51.2% accuracy). Surely, the standard will be different for different sorts of problems. When it comes to identifying a place to dig for gold, a prospector might be doing extremely well with a 30% hit rate. But when it comes to predicting whether Larry Bird will make the next free throw (where one knows he makes 90% of his free throws), a 60% hit rate would be atrocious. After all, one can get a 90% hit rate by always predicting he'll make it. Another way to tweak reliabilism is to recognize that for some problems, rationality might not require maximizing reliability but minimizing certain kinds of unreliability. For example, a rational predator detector would minimize false negatives (e.g., failing to detect the lion) even if the detector's sensitivity brought with it a somewhat higher false positive rate (e.g., false lion alarms). As far as we know, no reliabilist has yet offered a serious and detailed case for this sort of flexibility.

PRAGMATISM

Pragmatism, like reliabilism, is a consequentialist theory of rationality. It evaluates reasoning in terms of its consequences. This is very different from the Standard Picture, which takes certain rules (of logic or probability) to define rational thought. The reliabilist evaluates reasoning in terms of whether it has a settled tendency to deliver true beliefs, and the pragmatist evaluates reasoning in terms of whether it has a settled tendency to deliver useful beliefs. In recent years, pragmatism about rationality has been championed by Stephen Stich.[17,41] According to Stich, 'One system of cognitive mechanisms is preferable to another if, in using it, we are more likely to achieve those things that we intrinsically value' (Ref 17, p. 24).

> In evaluating systems of cognitive processes, the system to be preferred is the one that would be most likely to achieve those things that are intrinsically valued by the person whose interests are relevant to the purposes of the evaluation. In most cases, the relevant person will be the one who is or might be using the system. So, for example, if the issue at hand is the evaluation of Smith's system of cognitive processes in comparison with some actual or hypothetical alternative, the system that comes out higher on the pragmatist account of cognitive evaluation is the one that is most likely to lead to the things that Smith finds intriniscally valuable (Ref 17, p. 131–132).

Pragmatism's evaluations will depend on a reasoner's environment. This is a feature of any consequentialist view of rationality: a reasoning strategy might have good consequences in one environment and bad consequences in a different environment. Pragmatism's evaluations will also depend on what a reasoner intrinsically values. And this is perhaps a more disturbing implication. Pragmatism might offer strange prescriptions for a person with strange intrinsic values. But this is not an implication that is likely to bother the pragmatist. As Stich argues: 'Relativism in the evaluation of reasoning strategies is no more worrisome than relativism in the evaluation of diets or investment strategies or exercise programmes' (Ref 17, p. 9).

For two reasons, pragmatism can makes what counts as rational reasoning very difficult, or even practically impossible, to figure out. Begin with the assumption that the remote consequences of our beliefs (and so of our belief-forming mechanisms) can be very significant to our intrinsic values. For example, reasoning to the belief that *this* summer class will be more enjoyable than *that* summer class might make the difference between becoming an unhappy lawyer or a flourishing doctor. But such contingencies are unpredictable. Not all forms of consequentialism have this problem. Determining whether some cognitive mechanism produces mostly true beliefs can be in practice difficult. But it is usually a question of whether the mechanism tracks the right sorts of regularities in the local environment. The issue of the reliability of a cognitive mechanism is contained in a way the issue of its practical utility is not.

The second reason pragmatism tends to make rationality epistemically inaccessible has to do with the nature of a person's intrinsic values. The intuitive idea is clear enough: one intrinsically values whatever one values for its own sake. But one might intrinsically value a certain political ideal (economic freedom) or a character trait (total honesty) or another person (a spouse) on the basis of a misapprehension. If one were to be fully informed about the object she values, she would no longer find that ideal, trait or person valuable, intrinsically, or otherwise. Indeed, she might find that what she thought was valuable is actually quite harmful. This worry is captured by Oscar Wilde's witticism, 'When the gods wish to punish us, they answer our prayers.'

The obvious solution for the pragmatist is to argue that one intrinsically values whatever one would value for its own sake if appropriately informed. This solution, however, makes intrinsic value at least somewhat inaccessible: One's intrinsic values are not a simple function of those things one actually values for their own sake. There will be items one actually values for their own sake that one does not intrinsically value; and there will be items one intrinsically values that one does not actually value for their own sake. Just how inaccessible intrinsic value is will depend on the details of the theory. But getting useful advice from the pragmatist is going to be limited to the extent intrinsic value is inaccessible. It would be ironic if pragmatism were felled by its failure to deliver useful epistemological advice.

A natural criticism of pragmatism about rationality is that it confuses the epistemic with the pragmatic. The complaint might best be stated with some table pounding for emphasis, 'But the pragmatist doesn't provide an account of *epistemic* evaluation!' In reply to this objection, Stich has embraced 'the very Jamesian contention that *there are no intrinsic epistemic virtues*' (Ref 17, p. 24). 'For pragmatists, there are *no* special cognitive or epistemological values. There are just *values*' (Ref 41, p. 9). The pragmatist is a normative monist. Regardless of the item one is evaluating—an instance of reasoning, an action, a work of art, a corkscrew—the evaluative standard is the same: How likely is this to bring about those things one intrinsically values?

This normative monism is contrary to a more traditional normative pluralism. Tradition partitions the normative realm into various domains—the moral, the practical, the aesthetic, the epistemological. While we cannot hope to resolve this monism-pluralism dispute here, it is worth briefly setting the debate up in a bit more detail. The reason one might adopt the traditional, pluralistic view is that the different normative domains seem to represent real and distinct classes of problems that humans face: How can we live together in harmony? (Ethics. Political Philosophy.) How should one reason about the world? (Epistemology.) How should various groups of people come together to make judgments? (Social Epistemology.) How should one go about actively constructing a good life for oneself and those one cares about? (Practical Reason, Aesthetics, Political Philosophy.) Because these problems are different, *prima facie* it seems that the norms we apply to them are also different. For example, when someone says they believe that Barack Obama was born in Kenya, we criticize them for lacking good evidence; when someone injures another person, we criticize them for causing harm. So the prima facie challenge for the pragmatist, or any normative monist, is how to handle this apparent normative diversity in terms of a single evaluative standard.

CONCLUSION

This review has focused on the prospects and challenges of rule-based and consequentialist theories of rationality. These theories evaluate reasoning directly. They do not make rationality derivative on other epistemic categories. Some might argue that rationality is derivative on the notion of justification: rational reasoning leads to justified belief. Others might argue that it is derivative on the notion of a virtuous reasoner: rational reasoning is the product of the exercise of intellectual virtue. Our task here has been to explore theories that evaluate reasoning directly, without the intervention of prior epistemological categories. And while we perhaps have not resolved many issues, we have hopefully made questions about the nature of rationality clear and vivid.

REFERENCES

1. Nisbett RE, Borgida E. Attribution and the psychology of prediction. *J Pers Soc Psychol* 1975, 32:932–943, p. 935.
2. Kahneman PS, Tversky A. *Judgment Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press; 1982.

3. Casscells W, Schoenberger A, Graboys TB. Interpretation by physicians of clinical laboratory results. *N Engl J Med* 1978, 2:999–1001.
4. Gigerenzer G, Hell W, Blank H. Presentation and content: the use of base rates as a continuous variable. *J Exp Psychol Hum Percept Perform* 1988, 14:513–525.
5. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. *Psychol Rev* 1995, 102:684–704.
6. Meehl PE. *Clinical Versus Statistical Prediction A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: University of Minnesota Press; 1954.
7. Howard JW, Dawes RM. Linear prediction of marital happiness. *Pers Soc Psychol Bull* 1976, 2:478–480.
8. Carroll JS, Wiener RL, Coates D, Galegher J, Alibrio JJ. Evaluation, diagnosis, and predicition. *Law Soc Rev* 1982, 17:199–228.
9. DeVaul RA, Jervey F, Chappell JA, Carver P, Short B, O'Keefe S. Medical school performance of initially rejected students. *J Am Med Assoc* 1957, 257:47–51.
10. Swets JA, Dawes RM, Monahan J. Psychological science can improve diagnostic decisions. *Psychol Sci Public Interest* 2000, 1:1–26.
11. Dawes RM. A case study of graduate admissions: application of three principles of human decision making. *Am Psychol* 1971, 26:180–188.
12. Stillwell W. Evaluating credit applications: a validation of multiattribute utility weight elicitation techniques. *Organ Behav Hum Perform* 1983, 32:87–108.
13. Carpenter RG, Gardner A, McWeeny PM, Emery JL. Multistage scoring system for identifying infants at risk of unexpected death. *Arch Dis Child* 1977, 52:606–612.
14. Golding J, Limerick S, Macfarlane A. *Sudden Infant Death: Patterns, Puzzles, and Problems*. Seattle, WA: University of Washington Press; 1985.
15. Ashenfelter O. Predicting the quality and prices of Bordeaux wine. *Econ J* 2008, 118:174–184.
16. Gigerenzer G, Todd PM. *Simple Heuristics that Make Us Smart*. New York: Oxford University Press; 1999.
17. Stich SP. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: MIT Press; 1990.
18. Stanovich KE. *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates; 1999.
19. Nisbett RK, Peng IC, Norenzayan A. Culture and systems of thought: holistic vs. analytic cognition. *Psychol Rev* 2001, 108:291–310.
20. Nisbett RE. *The Geography of Thought: How Asians and Westerners Think Differently—and Why*. New York: Free Press; 2003.
21. Stich S. Philosophy and WEIRD intuition. *Behav Brain Sci* 2010, 33:110–111.
22. Stein E. *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science*. Oxford: Clarendon Press; 1996.
23. Piattelli-Palmarini M. *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*. New York: Wiley; 1994.
24. Bishop MA, Trout JD. *Epistemology and the Psychology of Human Judgment*. New York: Oxford University Press; 2005.
25. Cohen LJ. Can human irrationality be experimentally demonstrated? *Behav Brain Sci* 1981, 4:317–370.
26. Gigerenzer G. How to make cognitive illusions disappear: Beyond "heuristics and biases". *Eur Rev Soc Psychol* 1991, 2:83–115.
27. Kahneman D, Tversky A. On the reality of cognitive illusions: a reply to Gigerenzer's critique. *Psychol Rev* 1996, 103:582–591.
28. Gigerenzer G. On narrow norms and vague heuristics: a reply to Kahneman and Tversky. *Psychol Rev* 1996, 103:592–596.
29. Vranas PB. Gigerenzer's normative critique of Kahneman and Tversky. *Cognition* 2000, 76:179–193.
30. Sosa E, Galloway D. Man the rational animal? *Synthese* 2001, 122:165–178.
31. Nisbett RE. *Rules for Reasoning*. Hillsdale, NJ: L. Erlbaum Associates; 1993.
32. Talbot, W. (2001). Bayesian epistemology. In *Stanford Encyclopedia of Philosophy*. Available at: http://plato.stanford.edu/entries/epistemology-bayesian/ (Accessed October 12, 2012)
33. Jeffrey RC. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press; 1992.
34. Harman G. *Change in View: Principles of Reasoning*. Cambridge, MA: MIT Press; 1986.
35. Cherniak C. *Minimal Rationality*. Cambridge, MA: MIT Press; 1986.
36. Simon H. A behavioral model of rational choice. In: *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: John Wiley & Sons; 1957.
37. Gigerenzer G, Selten R. *Bounded Rationality*. Cambridge, MA: MIT Press; 2001.
38. Goldman A. Internalism exposed. *J Philos* 1999, 96:271–293.
39. Feldman R, Connee E. Internalism defended. *Am Philos Q* 2001, 38:1–18.
40. Kornblith H. *Knowledge and Its Place in Nature*. Oxford: Clarendon Press; 2002.
41. Stich S. Naturalizing epistemology: Quine, Simon and the prospects for pragmatism. *R Inst Philos Suppl* 1993, 34:1–17.

READINGS AND RESOURCES

Heuristics and Biases

Gilovich T, Griffin D, Kahneman D. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press; 2008.

Kahneman D, Slovic P, Tversky A. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press; 1982.

Kahneman D, Tversky A. *Choices, Values, and Frames*. Cambridge: Cambridge University Press; 2000.

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

Plous S. *The Psychology of Judgment and Decision Making*. Philadelphia, PA: Temple University Press; 1993.

Piattelli-Palmarini M. *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*. New York: John Wiley & Sons; 1994.

Predictive Modeling

Bishop M. In praise of epistemic irresponsibility: how lazy and ignorant can you be? *Synthese* 2000, 122:179–208.

Bishop MA, Trout JD. 50 years of successful predictive modeling should be enough: lessons for philosophy of science. *Philos Sci* 2002, 69(S3):S197–S208. doi: 10.1086/341846.

Dawes R. The robust beauty of improper linear models in decision making. In: Kahneman D, Slovic P, Tversky A, eds. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press; 1982, 391–407.

Dawes R. *House of Cards: Psychology and Psychotherapy Built on Myth*. Macmillan, New York: The Free Press; 1994.

Dawes R. A theory of rationality as a 'reasonable' response to an incomplete specification. *Synthese* 2000, 122: 133–163.

Dawes R, Corrigan B. Linear models in decision making. *Psychol Bull* 1974, 81:95–106.

Dawes R, Faust D, Meehl P. Clinical versus actuarial judgment. *Science* 1989, 243:1668–1674.

Grove WM, Meehl PE. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychol Public Policy Law* 1996, 2:293–323.

Fast and Frugal Heuristics

Bishop MA. Fast and frugal heuristics. *Philos Compass* 2006, 1:201–223.

Gigerenzer, G. (1991) On cognitive illusions and rationality. Poznan *Stud Philos Sci Humanit*, 21, 225-249.

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

Goldstein DG, Gigerenzer G. Models of ecological rationality: the recognition heuristic. *Psychol Rev* 2002, 109:75–90.

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

Todd PM, Gigerenzer G. *Ecological Rationality: Intelligence in the World*. Oxford: Oxford University Press; 2012.

Cognitive Diversity

Nisbett RE, Cohen D. *Culture of Honor: The Psychology of Violence in the South*. Boulder, CO: Westview Press; 1996.

Peng K, Nisbett RE. Culture, dialectics, and reasoning about contradiction. *Am Psychol* 1999, 54:741–754.

Stanovich KE, West RF. Individual differences in reasoning: implications for the rationality debate? *Behav Brain Sci* 2000, 23:645–726.

Reject the Norm Arguments

Cohen JL. On the psychology of prediction: whose is the fallacy? *Cognition* 1979, 7:385–407.

Kahneman D, Tversky A. On the psychology of prediction. *Psychol Rev* 1973, 80:237–251.

Kahneman D, Tversky A. On the interpretation of intuitive probability: a reply to Jonathan Cohen. *Cognition* 1979, 7:409–411.

Bayesianism

Earman J. *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press; 1992.

Goodman N. *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press; 1983.

Horwich P. *Probability and Evidence*. Cambridge: Cambridge University Press; 1982.

Jaynes ET(Larry Bretthorst G, ed.). *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press; 2003.

Jeffrey R. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press; 1992.

Sober E. Bayesianism—Its Scope and Limits. In: Swinburne R, ed. *Bayes's Theorem*. Oxford: Oxford University Press; 2002, 21–38.

Problems with the Standard Picture of Rationality

Goldman AI. *Justification and Reliability in Epistemology and Cognition*. Cambridge, MA: Harvard University Press; 1986.

Reliabilism about Rationality

Bishop MA, Trout JD. Strategic reliabilism: A naturalistic approach to epistemology. *Philosophy Compass* 2008, 3:1049–1065.

Goldman AI. *Epistemology and Cognition*. Cambridge, MA: Harvard University Press; 1986.

Kornblith H. Justified belief and epistemically responsible action. *Philos Rev* 1983, 92:33–48.

Kornblith H. The laws of thought. *Philos Phenomenol Res* 1992, 52:895–911.

Pragmatism

Bishop MA. Reflections on cognitive and epistemic diversity: can a Stich in time save Quine?. In: *Stich and His Critics*. Chichester: Wiley-Blackwell; 2009, 113–136.

Davidson D. Truth rehabilitated. In: *Truth, Language, and History*. Oxford: Clarendon Press; 2000, 3–18.

Dewey J. *The essential Dewey*, vol. 1 & 2. Bloomington: Indiana University Press; 1999(Hickman L and Alexander T, eds).

James W. *Principles of Psychology*. Cambridge, MA: Harvard University Press; 1890, 1981.

James W. *The Will to Believe and Other Popular Essays in Philosophy*. Cambridge MA: Harvard University Press; 1897.

James W. *Varieties of Religious Experience*. Cambridge MA: Harvard University Press; 1902.

James W. *Pragmatism: A New Name for Some Old Ways of Thinking*. Cambridge MA: Harvard University Press; 1907.

James W. *The Meaning of Truth*. Cambridge MA: Harvard University Press; 1909.

Peirce CS. *The Essential Peirce*, vol. 1. Bloomington: Indiana University Press; 1992(edited by the Peirce edition project).

Peirce CS. *The Essential Peirce*, vol. 2. Bloomington: Indiana University Press; 1999(edited by the Peirce edition project).

Rorty R. *Philosophy and the Mirror of Nature*. Oxford: Blackwell; 1980.

Rorty R. *The Consequences of Pragmatism*. Sussex: Harvester; 1982.

Stich SP, Cohen JL, Goldman AI, Harman G, Lycan WG. Book symposium: fragmentation of reason. *Philos Phenomenol Res* 1991, 51:179–213.