



# Debunking Biased Thinkers (Including Ourselves)

**ABSTRACT:** *Most of what we believe comes to us from the word of others, but we do not always believe what we are told. We often reject thinkers' reports by attributing biases to them. We may call this debunking. In this essay, I consider how debunking might work and then examine whether, and how often, it can help to preserve rational belief in the face of disagreement.*

**KEYWORDS:** epistemology, social epistemology, skepticism, sources of knowledge and justification, philosophy of psychology

Our eyes see nothing behind us. A hundred times a day we make fun of ourselves in the person of our neighbor and detest in others the defects that are more clearly in ourselves, and wonder at them with prodigious impudence and heedlessness.  
—Montaigne, 'Of the Art of Discussion'

Most of what we believe comes to us from the word of others. But we do not always believe what we are told. Sometimes we reject thinkers' reports by attributing biases to them. We often reason in that way when we disagree with what others tell us. Here are a few cases that capture the reasoning:

- You think that some climate science researchers have been biased by powerful financial incentives from the oil industry and so you regard their reports about climate change as not worth placing confidence in.
- A smart but hotheaded talk show host is biased by disdain for her opponents. We take ourselves to have reason to reject her statements about her opponents' views and personal lives.
- Hilary Putnam and Robert Nozick disagreed about political philosophy. In his 1982 book, *Reason, Truth and History*, Putnam tells us that he and Nozick had talked extensively about politics,

For helpful comments, I am grateful to Tomás Bogardus, William Dyer, Ian Evans, Chris Frey, Stephen Grimm, Michael Hannon, Xingming Hu, Peter King, Sam Kampa, Matthew Kisner, Charles Lassiter, Kirk Lougheed, Joshua Rasmussen, Andrew Rotondo, Angela Sager, Machessa Samz, Peter Seipel, Coran Stewart, Mark Timmons, and Brandon Warmke. Joe Morton and Justin Reppert helped me locate sources for footnote 9. Special thanks are due to Jennifer Ballantyne, E. J. Coffman, Bryan Frances, Nathan King, Shane Wilkins, Benjamin Wilson, and two anonymous referees for their generous help.



but—despite their patient, open-minded discussion—the dispute was never resolved. Putnam asked: ‘What happens in such disagreements? When they are intelligently conducted on both sides, sometimes all that can happen is that one sensitively diagnoses and delineates the source of the disagreement’ (1982: 164). Putnam’s diagnosis of Nozick’s error came to this: Nozick had a certain ‘complex of emotions and judgments’ that prevented him from enjoying a ‘certain kind of sensitivity and perception’ about politics (1982: 165). Putnam thought his colleague was subject to ‘powerful forces of a non-rational kind [that tend] to sway our judgment’ (1982: 167). Putnam didn’t change his political views in light of the disagreement at least in part because he regarded Nozick as biased.

In each case, someone attributes a bias to a disagreeing thinker and then downgrades the testimonial evidence the thinker has offered. The testimonial evidence may have had an impact on someone’s attitude—for instance, you may have changed your mind about climate change once you listened to the industry-funded researchers, or Putnam may have changed his political opinions in light of Nozick’s disagreement. But any downward push in confidence is diminished, at least to some extent, by debunking or discounting the testimonial evidence.

In recent discussion of epistemological questions concerning disagreement, the significance of this sort of reasoning has been noted. One central theoretical divide is over the following question: does learning that we disagree with an epistemic peer (i.e., someone roughly equally informed and competent to answer some question) *always* give us reason to reduce our confidence in our view? So-called conciliationists think it does, but nonconciliationists argue that some cases of recognized disagreement between peers allow at least one of them to retain confidence (see Feldman and Warfield [2010] and Christensen and Lackey [2013] for discussion). But both conciliationists and nonconciliationists agree that attributing biases to those who disagree is one important strategy for reacting to disagreements. Theorists in both camps think that if we have reason to regard another thinker as biased, we can sometimes demote that thinker from peer status. The thought that we can, and do, rationally attribute biases to others in order to prevent their dissent from lowering our confidence in our views is widely affirmed in discussions of disagreement.

Let us call this *debunking reasoning*. As far as I know, the reasoning has not been explored in detail. I hope to understand it better. To that end, I will address four main questions:

- What is debunking reasoning?
- What are good reasons to debunk testimony from thinkers who disagree with us?
- How do people tend to make judgments about biases?
- How often do we have good reasons to debunk testimony from thinkers who disagree with us?

Here is my overall thesis. Although debunking reasoning promises to preserve rational belief in the face of disagreement, debunking often won't be reasonable for us once we realize what it demands and once we know how people tend to make judgments about biases. Our discussion delivers two main lessons, which I'll return to at the conclusion. First, in the face of testimony from those who disagree with us, we may have considerable reason to become less confident in our views because we won't be able to debunk dissenting testimony as often as we might like. Second, we need principled methods for assessing bias in others and in ourselves, and this calls for a dramatic shift in how we make judgments about bias.

Before we begin, let me underscore why debunking testimonial evidence by attributing bias to the testifier is a topic worthy of our attention. Debunking plays a crucial practical role while we navigate our information-saturated world. On the one hand, we can't accept everything we hear—on pain of incoherence or the possibility of nonstop waffling in our views. On the other hand, if we 'fact-check' every last bit of testimony, we will be hopelessly bogged down in gathering further evidence to make up our minds. Judging that an evidence source is biased lets us swiftly judge its fruits as bad and move on. And reasoning in this way helps us piece together and manage our webs of belief. It is one type of inferential method among others that we use to build our intellectual houses. My own view is that the methods by which we amass and evaluate evidence on difficult, non-obvious matters are often makeshift. Doesn't it occasionally seem that our intellectual houses are poorly constructed, that certain walls or rooms would be rent to chips and splinters if they were truly tested? This will be no surprise if our methods are makeshift, as I say they are. To improve our intellectual houses, we must first scrutinize our building methods.

## I. Debunking Reasoning

Debunking reasoning plays a role in recent debates over the epistemic significance of disagreement. Everyone agrees that learning of disagreement is evidence that may undermine our views. And everyone agrees that reasonably attributing biases to disagreeing thinkers lets us resist a downward push in our confidence in our views. But conciliationists and nonconciliationists have different views about when debunking is appropriate.

For example, David Christensen, a conciliationist, thinks that recognized peer disagreement always provides a reason to reduce confidence in one's own views substantially, unless one has an *independent reason* to downgrade the testimonial evidence one has received. An independent reason must not depend on the original reasoning used to reach one's own view (2007, 2011). This 'independence' constraint rules out reasoning of the following form: 'Well, McCoy disagrees with me about  $p$ . But  $p$  is true. So McCoy is wrong. And so I do not need to take her disagreement as a reason to doubt my view.' But an independent reason to think a peer is biased lets us debunk the peer's testimonial evidence, thinks Christensen (2014).

But debunking is not the exclusive practice of conciliationists. Nonconciliationists Michael Bergmann and Richard Fumerton have also suggested that we

can downgrade a disagreeing peer's testimony if we have reason to regard her as tending to err. Bergmann's idea is that if a thinker has in hand an explanation for why the peer has made an error—because the peer is biased in this particular case, for instance—then the explanation can preserve rational belief in the face of disagreement (2009). And here is what Fumerton says:

Do I have reason to suspect that some of my [disagreeing] colleagues are plagued by . . . subtle defects? Perhaps I have some reason to believe, for example, that they are the victims of various biases that cause them to believe what they want to believe. . . . Indeed, I suspect that I *do* have reason to believe that others are afflicted in such ways. (2010: 102)

By taking others to be biased, Fumerton thinks he may 'sometimes discount to some extent the fact that well-known and respected intellectuals disagree with me' (2010: 103).

On all sides, then, theorists allow reacting to disagreement by making bias attributions.<sup>1</sup> But how does the debunking reasoning work? Minimally, it involves two claims:

*Bias Premise:* Some thinker is biased regarding proposition *p*.

*Debunking Conclusion:* We have reason to reduce confidence in that thinker's report about *p*.

Any instance of debunking reasoning aims to secure the Debunking Conclusion at least partly on the basis of the Bias Premise. There's no one-size-fits-all argument schema to capture every path from premise to conclusion. That is because there is no single type of inference pattern here—it may be deductive, inductive, or abductive. Does this mean that we can't begin to understand, in general, when the debunking reasoning is successful and when it's not? No. We may safely assume that we can move in appropriate steps from the Bias Premise to the Debunking Conclusion. For instance, the following offers the makings of a deductively valid route from Bias Premise to Debunking Conclusion: we have evidence for the Bias Premise; evidence that a thinker is biased is evidence that the thinker is unreliable; and evidence that the thinker is unreliable makes it rational to reduce confidence in what the thinker reports or believes. But spelling out the reasoning in detail will not answer the important question about what counts as a good reason to accept the Bias Premise. That's where the action is.

<sup>1</sup> Let me note two other ways biases figure in discussions of disagreement. Bias-attributions come up in work on disagreement with experts (e.g., Frances 2013: 128–29). I focus on peer disagreement here, but my proposals bear on debunking in other types of disagreements, too. Second, as an anonymous referee pointed out to me, learning that testifiers have been selected in a biased way may give us reason to discount their reports. For example, a wise juror knows an attorney will dredge up experts for her side of the case; the juror may sometimes downgrade an expert's testimony due to a biased selection procedure even without reason to think the expert is subject to psychological bias.

## 2. Debunking Strategies

So what makes it reasonable to accept the Bias Premise? As we work toward an answer, notice what the Bias Premise is. It is a claim about a thinker's disposition, in some situation, to form inaccurate or unreasonable attitudes regarding a proposition  $p$ . Biases may prevent someone from accurately forming attitudes about  $p$  or from being reasonable in responding to evidence regarding  $p$ . Being biased doesn't imply actual inaccuracy or unreasonableness, but it carries a nontrivial risk of error, due to a tendency to form attitudes or evaluate evidence in an unreliable way. Being biased means being unreliable to some degree or other, but the converse doesn't hold. Being unreliable does not mean being biased—the victim of a Cartesian demon is unreliable but not biased, for instance. Thus, affirming the Bias Premise calls for evidence that tells us about a thinker's attitude-forming tendencies.

Let us try out four potential strategies for securing the Bias Premise. I will begin with a strategy that may tempt us. Suppose we disagree with a peer and have reason to think our view is correct. But if our view is correct, then our peer's view is mistaken. It follows that we have reason to regard our peer as mistaken. Now, in peer disagreements, we can rule out certain explanations for a peer's mistake—it's not for lack of relevant evidence or intellectual abilities that a peer is wrong. But abilities are a kind of competence, and even the competent may sometimes blow it. Let's also assume, then, that we can't attribute our peer's mistake to a mere performance error. It follows that the peer's mistake is due to bias. Thus, the Bias Premise is true.

We can summarize that reasoning as follows:

*Dogmatic Dismissal Strategy.* We have reasons to accept (1) that the attitude we take toward proposition  $p$  is correct, (2) if we are correct about  $p$ , then a disagreeing thinker's attitude regarding  $p$  is mistaken, (3) the disagreeing thinker's mistake is not due to a lack of relevant evidence or intellectual abilities or to a performance error, and (4) if a thinker's mistake is not explained by a lack of relevant evidence and/or intellectual abilities or by a performance error, it is (probably) due to bias. On the basis of (1) through (4), and what follows from those steps, we can infer that the Bias Premise is (probably) true.

On a first look, that may seem to be a promising route to the Bias Premise. But the Dogmatic Dismissal Strategy is problematic.

Here is a crucial step in the reasoning: our view is correct, and if we are correct, then our opponent is mistaken, and so our opponent's view is mistaken. But this reasoning, if it's any good, licenses an extremely dogmatic response to any and all opponents. Whenever others disagree with us—even recognized epistemic *superiors*—this strategy lets us infer that they're mistaken by appealing to our reasons for thinking that we are correct. If we have such reasons in the first place, then we need never change our minds when encountering disagreement. Something has gone wrong.

Consider a better strategy. Suppose that we have some evidence—call it *E*—and it is reasonable for us to think that having *E* *rationally compels* some attitude toward a proposition. Now if we know that a thinker has *E* but does not hold the attitude mandated by *E*, then we can conclude that the thinker is in error. To illustrate this, consider an example from Peter van Inwagen:

There exists an organization called the Flat Earth Society, which is, as one might have guessed, devoted to defending the thesis that the earth is flat. At least some of the members of this society are very clever and are fully aware of the data and arguments—including photographs taken from space—that establish that the earth is spherical. Apparently this is not a joke; they seem to be quite sincere. What can we say about them except that they are intellectually perverse? (2009: 21)

Van Inwagen's idea is that members of the Flat Earth Society are in an evidential situation that rationally compels them to accept that the earth is spherical. It's an established fact that the world is not flat, thinks van Inwagen, because there is compelling, knockdown evidence widely available for that thesis. Since van Inwagen knows these people nevertheless believe the world is flat, he concludes that they are 'intellectually perverse'. Let us assume that the sort of intellectual perversity at issue here involves bias. All of this suggests a second route to the Bias Premise:

*Unresponsiveness-to-Compelling-Evidence Strategy.* We have reasons to accept (1) that some body of evidence *E* rationally compels a particular attitude toward a proposition *p*, (2) that a disagreeing thinker has *E* but does not hold the attitude toward *p* required by *E*, and (3) that the best explanation for why the disagreeing thinker does not hold the required attitude toward *p* is bias. On the basis of (1) through (3), we can infer that the Bias Premise is (probably) true.

In the disagreements we are party to, do we plausibly have evidence that rationally compels a particular attitude? If we do not, this strategy won't help.

Let's begin close to home and reflect for a moment on philosophical disagreements. Could we use the Unresponsiveness-to-Compelling-Evidence Strategy to debunk fellow philosophers? It's hard to say, partly because it is doubtful whether there are any philosophical arguments that rationally compel a particular attitude. Philosophers like van Inwagen (2009: 34, 105) and David Lewis (1983: x) deny that there are conclusive, knockdown arguments in our field, at least for substantive theses. But even philosophers who fancy themselves to have discovered knockdown arguments have failed to convince everyone or even a majority of philosophers. The uncertain among us therefore may find it unclear whether or not there are any compelling philosophical arguments.<sup>2</sup>

<sup>2</sup> For discussion of whether there are conclusive arguments in philosophy, see Ballantyne (2014) and Kelly and McGrath (forthcoming).

Even if there are compelling arguments for only a few controversial philosophical positions, it's plausible that some nonphilosophical claims are supported by compelling evidence. For instance, van Inwagen's case of the Flat Earth Society is like that. So is the case of scientists funded by the oil industry who deny certain claims about human-caused climate change. In such cases, the available evidence seems to demand a particular doxastic response. Since certain people have that evidence and yet fail to hold the mandated views, we can conclude they are biased. But in most of the disagreements we care about we won't have reason to think our evidence is so potent. Two of the cases that we began with—the ones featuring Putnam and Nozick and the smart but hotheaded talk show host—do not seem to involve compelling, knockdown evidence for a thesis. At best, the Unresponsiveness-to-Compelling-Evidence Strategy will have limited use. We can leave it to the side.

Let's turn now to a strategy for reaching the Bias Premise that promises to apply more widely, starting with an example offered by David Christensen:

[You are] attending a contest for high-school musicians. After hearing all the performers, and becoming quite convinced that Kirsten's performance was significantly better than Aksel's, [you] hear the man next to [you] express just the opposite opinion. But then [you] find out that he is Aksel's father. (2014: 143)

Christensen thinks you can properly downgrade Aksel's father's testimony. More important, Christensen denies that anything along the lines of the Dogmatic Dismissal Strategy will explain why. According to Christensen, a good reason for reducing confidence in Aksel's father's testimony is *not* that you are sure that Kirsten's performance was significantly better than Aksel's. If that were your reason, of course, you would not have an 'independent' reason to debunk the testimony.

What would be a better reason then? You attribute to Aksel's father a bias with respect to his judgment concerning the relative merits of the musical performances. Why think that Aksel's father is biased? Most of us can't be expected to judge our kinfolk impartially on the merits of their musical performances. Family relationships bias our judgments here as elsewhere. Thus, you can reasonably accept the Bias Premise in this case and reach the Debunking Conclusion.

But it is doubtful that all of this captures a sufficient condition for accepting the Bias Premise. Imagine a case just like Christensen's, with one extra detail: you know a further factor holds for Aksel's father, a factor that tends to counteract or neutralize the biasing factor. More specifically, imagine that Aksel's father is a professional judge on the high-school music-contest circuit and that he has special training that helps him avoid biases toward his own students. Aksel's father may be *enlightened* by his music-contest-judgment expertise, and perhaps this lets him avoid or overcome the biasing influence of fatherhood. In this situation, you would not have reason to infer that Aksel's father is biased. Reason to think fatherhood biases judgments and that Aksel's father enjoys paternity, is thus not enough to accept the Bias Premise.

Putting that all together, here is what I'll call the

*Biasing-Factor-Attribution Strategy.* We have reasons to accept (1) that a factor *F* tends to bias judgments about proposition *p*, (2) that factor *F* holds for a disagreeing thinker's judgment about *p*, and (3) that we know of no 'enlightening' factor that holds for the thinker's judgment about *p* (i.e., a factor that tends to counteract or neutralize *F*'s influence). On the basis of (1) through (3), we infer that the Bias Premise is (probably) true.

Here we find an inference that's a kind of statistical syllogism. It appears to be available in a wide range of cases as there are many factors that bias thought. I'll examine this strategy in some detail below after we've looked at how people form judgments about bias, but for now let's move on to a different strategy.

As I noted earlier, Richard Fumerton takes himself to have reason to regard disagreeing thinkers as suffering from biases. What is his reason? He writes:

Do I have reason to suspect that some of my [disagreeing] colleagues are plagued by . . . subtle defects? Perhaps I have some reason to believe, for example, that they are the victims of various biases that cause them to believe what they want to believe. . . . Indeed, I suspect that I *do* have reason to believe that others are afflicted in such ways. . . . I do, in fact, think that I have got more self-knowledge than a great many other academics I know, and I think that self-knowledge gives me a better and more neutral perspective on a host of philosophical and political issues. I suspect that it is in part the fact that I take this belief of mine to be justified that I do think that I can sometimes discount to some extent the fact that well-known and respected intellectuals disagree with me.  
(2010: 102–3)

Fumerton doesn't rely on the Biasing-Factor-Attribution Strategy. He doesn't point to some biasing factor that holds for his opponents. So how does Fumerton's reasoning deliver the Bias Premise? Doesn't it presuppose that disagreeing thinkers are biased?

Squinting a bit at what Fumerton says, his reasoning may proceed as follows. Unlike the Biasing-Factor-Attribution Strategy, which involves a kind of statistical syllogism, Fumerton's reasoning relies on an inference to the best explanation. Fumerton takes himself to know things about himself that indicate he is not biased or that at least he is relatively less biased than thinkers he disagrees with on politics and philosophy. But since these disputes (presumably) concern matters of objective fact, someone has made a mistake. Since Fumerton assumes that he and his opponents share relevant evidence and intellectual abilities, the best way to explain one side's error is to posit a bias. Given that Fumerton purportedly knows he's neutral, it's his opponents, then, who must be biased. Thus, Fumerton has a reason to accept that his opponents are biased.



We may sum up this reasoning as follows:

*Self-Exculpating Strategy.* We have reasons to accept (1) that we are not biased and (2) that one side of the disagreement has made a mistake due to bias (rather than differences in evidence or intellectual abilities or other factors including a performance error). On the basis of (1) and (2), we can infer that the Bias Premise is (probably) true.

I will consider this strategy below, but here's one remark on (2). Many disagreements can be explained without appealing to biases: differences in evidence or intellectual abilities as well as mere performance errors are included here. Therefore, reason to accept (2) sensibly amounts to reason to think alternative, nonbias explanations for why one side has erred are not as plausible as the explanation that bias offers.<sup>3</sup>

So far I've noted four reasons for accepting the Bias Premise. Since the Dogmatic Dismissal Strategy doesn't appear to offer us a good reason, we will ignore it. Although the Unresponsiveness-to-Compelling Evidence Strategy is sometimes effective, I'll set it aside, too, because its use is rather limited. The two remaining options—the Biasing-Factor-Attribution and the Self-Exculpating strategies—look more promising. Unfortunately, as I will argue, in a great many of the cases of disagreement we take to be practically or theoretically important those two debunking strategies won't help us out. But first let us turn to a third question: how do people tend to make judgments about biases?

### 3. The Bias Blind Spot

Psychologists have begun to reveal how we make judgments about biases, and the lessons are fascinating. Once we understand how people tend to form judgments about biases, we'll be able to assess critically the Biasing-Factor-Attribution and the Self-Exculpating strategies. Even if those strategies are sometimes rationally appropriate, we may often be unable to put them to use properly, given our knowledge of the psychological findings.

A growing body of work in psychology observes 'a broad and pervasive tendency for people to see the existence and operation of bias much more in others than in themselves' (Pronin 2007: 37). This is a kind of 'bias bias'—a bias that sways judgment and reasoning *about bias*—and it has been called the 'bias blind spot'. It results in the conviction that one's own judgments are less susceptible to bias

<sup>3</sup> To reasonably accept (2), we must rule out potentially minor asymmetries in relevant epistemic factors, but this may be difficult. For example, given that 'small' differences in evidence may sometimes make 'big' differences in how we should think, to accept (2) we'll sometimes need to have a reason to think that there's not even a small evidential difference. (King [2012] presses this point in connection with evidential symmetry between peers.) But we may assume that reasonably accepting (2) is not too hard, since we're supposing that many of us commonly attribute biases to recognized disagreeing peers. Without that assumption, the range of disagreements in which we can deploy the Self-Exculpating Strategy will be somewhat limited.

than the judgments of others. Direct testing confirms the blind spot is widespread (Pronin, Lin, and Ross 2002). Several cognitive mechanisms have been found to generate the bias blind spot: (1) an important evidential asymmetry between judgments of self and others, (2) the tendency to assume what has been called ‘naïve realism’, and (3) the motive of self-enhancement. I’ll explain each in turn.

When people make judgments about bias in themselves, they tend to rely on *introspective* evidence, but when they judge bias in others they tend to rely on *behavioral* or *extrospective* evidence. People look into their own minds to judge themselves but look at outward behavior when judging others, and this evidential asymmetry shapes judgment about bias.

But a central idea in psychology is that most biases are not reliably detected by introspection (Nisbett and Wilson 1977; Wilson and Brekke 1994; Kahneman 2003). We typically can’t figure out whether we are biased by merely gazing into our minds. Biases typically ‘leave no trace’ in consciousness. As Timothy Wilson and Nancy Brekke quip, ‘Human judgments—even very bad ones—do not smell’ (1994: 121). From the inside, biased attitudes seem just like unbiased ones. We can introspect the particular judgments our cognitive processes produce, but not the properties of the processes relevant to the judgments’ reliability. In other words, we can’t normally introspect the operation of biases on our judgments even when we can detect the outputs of those biases. The machine chugs along, but we can’t peek in to see whether it works properly.<sup>4</sup>

Although introspection does not reliably detect the operation of biases, that isn’t to say people don’t still try. But using introspection to attempt to discover bias has an important upshot: people may end up with the impression they have acted *in spite of* their own expectations and interests rather than *because of* them. As Joyce Ehrlinger, Thomas Gilovich, and Lee Ross point out, ‘one’s conscious efforts to have avoided bias, in fact to have “bent over backwards” to do so, are likely to be highly salient’ in one’s thinking about whether one is biased (2005: 686). The feeling that we’ve done our level best to be unbiased will encourage us to think we are unbiased, but that feeling should not be trusted.

Different judgments arise from differences in introspective access—no news there. But it gets worse. It’s not just that people actually rely asymmetrically on introspective and behavioral evidence. They also think they *should* place less weight on behavioral evidence in their own case and more weight on their own introspective evidence. In fact, researchers have noted that subjects sometimes show surprising disregard for the subjects’ own actions even when those actions become salient. Subjects insist their behavior isn’t relevant to decide whether they are biased. This is striking. So often in life we are judged by our actions, not our intentions or hopes or feelings. But in self-judgment about bias, we overlook our actions and instead cling to how things feel on the inside (Pronin and Kugler 2007).

<sup>4</sup> Can we ever introspect the judgment-reliability-relevant properties of biasing processes? Ehrlinger, Gilovich, and Ross note that ‘there are some occasions in which introspection does yield evidence of bias, or at least evidence of the motives that prompt bias. . . . A Little League umpire is apt both to realize and recollect how hard it was to utter “Strike three!” when it was his or her own child in the batter’s box’ (2005: 686).

The evidential asymmetry driving the blind spot seems to prove the old wisdom: nobody should be a judge in his own case (*nemo iudex in causa sua*).<sup>5</sup> But what if we break down the evidential asymmetry and inform people about others' thoughts and feelings? Will the blind spot vanish? Regrettably not. Even after people are given others' introspective reports, the blind spot tends to persist. In some experiments, psychologists provided observers with a sample of actors' introspections, but observers continued to impute more bias to others than to themselves even when the observers had read and listened to detailed reports from actors and believed those reports correctly reflected the actors' thoughts (Pronin and Kugler 2007).

A second source of the blind spot is our tendency to presume what psychologists have called 'naïve realism'—the idea that our experience of the world, others, and ourselves is veridical. We normally assume that our experience gives us a more or less 'unmediated' picture of how things really are.

How does naïve realism influence judgment about biases? It turns out that when we discover that others disagree, we often attribute biases to them. In one experiment, American and Canadian undergraduate students who disagreed with the US president's decision to invade Iraq attributed a greater degree of self-interest bias to the president than did students who agreed with him (Reeder et al. 2005). Those in the grips of naïve realism believe objective thinkers will agree with them. When others disagree with us, we are prompted to ask whether they've missed relevant evidence. If we think they are well-informed, naïve realism leads us to conclude they are biased (Pronin 2007: 39–40). Our tendency to try to resolve cognitive dissonance plausibly explains this sort of effect sometimes. When we find that people disagree, we treat this as *prima facie* evidence against our views. Then we resolve the dissonance in favor of our own views, often by way of bias attributions.

Evidential asymmetry and naïve realism are two sources of the bias blind spot: the motive of self-enhancement is a third. Research has established that we see ourselves in an overly positive light.<sup>6</sup> For valuable or desirable traits, we tend to overrate ourselves, even when the evidence suggests otherwise. In a classic study that should be close to every college teacher's heart, 94 percent of college teachers rated themselves as doing above-average work (Cross 1977). And when people lack a talent or positive trait, they are sometimes oblivious. These sorts of effects stem from powerful 'ego-protective' biases. We think well of ourselves, objective evidence be damned, but most people rarely notice this. In fact, self-enhancement bias is regarded as a key element of health. People with major depression tend to

<sup>5</sup> In fact, as an anonymous referee told me, some legal systems have maintained that no defendant should testify in her own case. The rationale was that her testimony would be biased. This was true in the United States during the nineteenth century, for example, when Susan B. Anthony couldn't testify against the charge that she voted illegally.

<sup>6</sup> Before the research was in, it had been anticipated by some observers of human nature. 'It is a law of our being', wrote Bertrand Russell, 'that, whenever it is in any way possible, we adopt beliefs as will preserve our self-respect' ([1928] 2004: 51). In the *Port-Royal Logic*, Antoine Arnauld and Pierre Nicole delivered a serious buzzkill: 'The human mind is not only naturally enamored of itself, but it is also naturally jealous, envious, and malicious toward others. It allows others to have advantages only with difficulty, because it desires them all for itself' ([1683] 1996: 207).

have more accurate self-assessments and fewer illusions about themselves, but at the cost of healthy functioning (Taylor and Brown 1988).

The self-enhancement motive gives us no protective illusions concerning other people, naturally enough (leaving aside, for now, the interesting case of family and friends). We attribute biases to others with ease. We expect them to make judgments that serve their self-interest and to make overly positive self-assessments. As a result, judgments of bias in self and others will differ. The motive to self-enhance thus makes us less likely to find bias in ourselves than in others (Pronin 2007: 37).

To sum up: the bias blind spot stems from an important evidential asymmetry, the assumption of naïve realism, and self-enhancement bias. These psychological mechanisms lead to the conviction that our judgments are less susceptible to bias than the judgments of others.

Here is an important part of the story we shouldn't ignore. To judge bias in others we focus on behavioral evidence, but that evidence by itself does not disclose others' inner dispositions. The evidence needs interpretation. We use 'abstract theories' of bias to decipher it. Let me sketch the content of our abstract theories (cf. Ross, Ehrlinger, and Gilovich, forthcoming). We know people are motivated to seek pleasure and avoid pain. They give heavy weight to their own wants and needs. They are sheltered from reality by various psychological defense mechanisms. People sometimes view issues through the lens of their ideology and social position. And they're capable of (self-)deception concerning the influence of nonrational factors on their judgments. These sorts of thoughts comprise our abstract theories. Our theories tell us when motives, needs, expectations, and context invite bias, allowing us to regard others' behavior as indicating the presence or risk of bias. These theories are rough guidelines. As a matter of fact, they are imperfect guides. For instance, experiments show that people sometimes cynically expect others to be more biased than they in fact are (Kruger and Gilovich 1999).

Thus concludes my review of psychological research on the blind spot. Let's see how this psychological picture might guide us in the business of debunking.

#### 4. Two Debunking Strategies and the Psychological Picture

Here is my fourth and final question. How often do we have good reasons to debunk testimony from those who disagree with us?

Recall that debunking reasoning moves from the Bias Premise to the Debunking Conclusion, and we earlier looked at four strategies to get the Bias Premise in hand. The Dogmatic Dismissal Strategy is a nonstarter. The Unresponsiveness-to-Compelling Evidence Strategy won't apply widely enough to situations where we hope to resist the downward push of testimony from disagreeing thinkers. But the Biasing-Factor-Attribution and the Self-Exculpating strategies show more promise. Let us consider, then, how often we can reasonably follow these two routes to the Bias Premise.

Awareness of the psychological picture brings trouble for anyone who hopes to deploy either the Biasing-Factor-Attribution or the Self-Exculpating strategies in a wide range of important cases. The trouble is that those debunking strategies require

particular judgments about biases, but reflection on the psychological picture raises doubts about whether we *reliably* make those judgments. Accordingly, we should often doubt that it's reasonable for us to deploy these strategies.<sup>7</sup>

Let us begin with the

*Self-Exculpating Strategy.* We have reasons to accept (1) that we are not biased and (2) that one side of the disagreement has made a mistake due to bias (rather than differences in evidence or intellectual abilities or other factors). On the basis of (1) and (2), we can infer that the Bias Premise is (probably) true.

As already noted, (2) requires reason to think alternative, nonbias explanations for why one side has erred are less plausible than the explanation bias provides. Let's grant for now that (2) can be reasonably accepted. I will argue that our inclination to accept (1) should be curbed by what we know about the bias blind spot.

Richard Fumerton appears to endorse the Self-Exculpating Strategy. Consider what he says about his reasoning: 'When I argue this way, I again risk sounding like a bit of a jerk. Do I really suppose that I am justified in thinking that there is an asymmetry between myself and others when it comes to various epistemic defects? Am I any less likely to be blinded to what is reasonable to believe by antecedent views or desires? Well, to be honest I suppose that I think I am' (2010: 105). How should we react to this disclosure? You may decide that Fumerton is a poster child for a public service ad campaign by psychologists to raise public awareness of the bias blind spot. ('The Bias Blind Spot: It Makes You Sound like a Jerk'.) But perhaps Fumerton's honesty is to be admired. The gist of his reasoning is pretty much standard operating procedure for us human beings, but none of this is normally expressed so candidly.

The blind spot neatly explains why Fumerton judges that (1) is true. He doesn't tell us his reasons for (1)—he just says he thinks he is justified to think it is true in some disagreements. Plausibly, Fumerton introspects to check for bias in himself and relies on behavioral evidence, guided by abstract theories of bias, to check for bias in others. Of course, introspective evidence is not a reliable way to recognize subconscious biases in ourselves and may even leave us with the feeling of having successfully overcome judgment-distorting influences after 'bending over backward' to be neutral.

Even so, we don't know that Fumerton is actually subject to the bias blind spot. All we know is what the psychological evidence says: humans in general tend to make judgments like his. An important issue, then, concerns the connection between the psychological evidence and a thinker's reasons to accept (1). Here's what I propose. Evidence about how people judge bias is relevant for assessing the premise that some person is not biased. The psychological evidence is the kind of evidence that tells us about how effectively we tend to assess evidence about bias. This is what has been called 'higher-order evidence'—the kind of evidence

<sup>7</sup>This leaves open the possibility that thinkers unaware of the psychological picture may commonly reasonably deploy the debunking strategies (cf. Kelly 2008: 629–30).

that tells us about our evaluation of, or responsiveness to, a body of first-order evidence (see Feldman 2005; Christensen 2010). Learning of the psychological evidence, then, tends to undermine belief in (1).<sup>8</sup> If we ourselves wish to accept (1), in full awareness of the psychological picture, we need to have a reason to think that we are not subject to the blind spot. Insofar as we lack reason to think we're not subject to it, we have reason to doubt (1) is true, and accordingly the Self-Exculpating Strategy will fail to deliver the Bias Premise.

An analogy will amplify the reasoning I have proposed (see Elga ms.). Imagine that Earhart is piloting a small aircraft above 10,000 feet. Earhart knows that people in small aircraft flying at high altitudes often suffer from hypoxia—a condition where the brain is oxygen-deprived—and, as a result, their judgments become unreliable. Once hypoxia has taken effect, it will typically seem to the hypoxic subject that her reasoning is perfectly good even when it is bad. Hypoxia has an effect on thinking without 'leaving a trace' in consciousness.<sup>9</sup> Since Earhart recognizes she may be hypoxic at this high altitude, she has reason to invest much less confidence in her calculations about her remaining fuel. To restore confidence in her calculations, she needs reason to think that she is not hypoxic.<sup>10</sup>

If her belief is reasonable, she needs independent means to determine that she's not hypoxic, such as an O<sub>2</sub> detector. And our situation with respect to bias is not unlike that of the high-flying pilot. Since we know about the bias blind spot, we should reduce confidence in (1), the premise that we are not biased, unless we have good reason to think that we are not suffering from the blind spot.

What counts as good reason to think we've overcome the blind spot? Back in the hypoxia example, once Earhart suspects she's hypoxic, a reason to regard herself as not hypoxic must arise from a belief-forming method she reasonably thinks is

<sup>8</sup> There are a priori arguments for the thesis that we can't regard one of our occurrent beliefs as biased (Winters 1979; Sorensen 1988: 104–7). My arguments remain neutral on that thesis. I only assume that our occurrent beliefs may properly change when we learn of the psychological evidence. If the a priori thesis is correct, the psychological evidence will never prompt us to believe that our belief is biased but must instead lead us to think a former belief was biased. But the a priori thesis, which Sorensen notes goes back to Kant, partly anticipates the blind spot research in a curious way. The thesis says we can't self-attribute bias, just as psychologists discovered we're not prone to do.

<sup>9</sup> To whom do we owe the pleasure of this knowledge? Intrepid scientists. For instance, J. S. Haldane and collaborators, including his son J. B. S., set up a steel chamber in a lab and some of them climbed in. A pump removed the chamber's air, and once the pressure equaled pressure at a high altitude, it was held steady using a tap. Test subjects kept notes to record their experience. In low oxygen conditions, their ideas grew strange and irrational, though they didn't know it until later. Haldane commented on his self-experimentation: 'I have always been quite confident in my own sanity, and it was only afterwards that I realized that I could not have been in a sane state of mind' (1922: 126). In one test, two scientists had planned to remain in the chamber for an hour at the lowest pressure possible 'without very serious impairment'. Their plan was foiled when onlookers noticed the two subjects were blue, very shaky, and could hardly stand up. At that point, 'the emergency tap was therefore opened so as to raise the pressure. There is a corresponding indignant and just legible note "some bastard has turned tap", after which the notes become quite legible again as the pressure rose' (1919: 185). For J. B. S. Haldane's charming description of one hypoxia self-experiment, see Haldane (2002: 70–71). I'm grateful to an anonymous referee for bringing J. B. S.'s essay to my attention.

<sup>10</sup> Suppose Earhart's belief that hypoxia makes her unreliable remains reasonable. If that belief is not reasonable, her confidence in her calculations may be restored—without her having reason to think she's not hypoxic.

unaffected by high altitude. Likewise, once we suspect we may be blinkered by the blind spot, any reason to think that we are unbiased must trace back to a method we reasonably think is not biased. That's why introspection alone won't normally dispel doubts about the blind spot—we know introspection is prejudiced in our favor.

If introspection is off-limits, where shall we turn? To avoid biases, thinkers sometimes try to *debias*—to identify and avoid biases or adjust their judgments to counteract the negative effects of biases (Wilson and Brekke 1994; Wilson, Centerbar, and Brekke 2002; Larrick 2004). Let's say we know of a reliable method to judge our susceptibility to the blind spot. If we reasonably think that our judgment that we aren't afflicted by the blind spot traces back to a reliable debiasing method, we get reason for thinking that we have avoided or overcome that bias. Thus, recognizing we've debiased would permit us to accept premise (1), our awareness of the blind spot notwithstanding.

As I'll now argue, reason to think that we've successfully debiased is going to be uncommon. That's because debiasing is extraordinarily hard. For a start, we naturally rely on abstract theories of bias to debias, but our theories may lead us astray. For instance, the feeling that we can debias by *carefully thinking things over* is tempting but mistaken (cf. Kenyon 2014: sec. 2). And so we'll want to take our cues from research on debiasing. But even when we do, debiasing in real life demands extensive knowledge of ourselves and the nature of our biases.

To see why, let's consider Timothy Wilson and Nancy Brekke's prominent account of debiasing (1994 and Wilson, Centerbar, and Brekke 2002). In brief, Wilson and Brekke see a thinker's inability to debias as stemming from a number of common sources: (1) that thinker's lack of awareness of his or her mental processes (e.g., the extent to which the thinker's positive evaluation was due to a self-enhancement motive), (2) lack of control over mental processes (e.g., the thinker's inability to prevent the fact that his or her status is at issue from influencing self-judgment), (3) inaccurate theories about biasing influences on judgment (e.g., the thinker's failure to appreciate how his or her own status could nonconsciously influence self-judgment), and (4) inadequate motivation to correct for bias (e.g., an insufficient desire to avoid a self-enhancing judgment) (cf. 2002: 187). Any one of (1) through (4) will prevent successful debiasing. In light of how commonly people find themselves in those conditions, Wilson and Brekke are 'rather pessimistic' about our ability to debias—and their pessimism even extends to trained psychologists who are familiar with the literature on biases (1994: 120; Wilson, Centerbar, and Brekke 2002: 190–91, 200). This bleak assessment confirms what other psychologists have found: debiasing demands knowledge individuals often lack (Kahneman 2003).

Psychologists say that our best shot at successful debiasing lies in debiasing techniques. One self-administered technique is to 'consider the opposite', to argue against one's own initial judgment.<sup>11</sup> This debiasing advice comes with a warning attached: this technique may backfire. 'Ironically, the more people try to consider

<sup>11</sup> See Larrick (2004: 323–24) and Wilson, Centerbar, and Brekke (2002: 197–98). Some debiasing techniques are launched at the social or organizational level: see Heath, Larrick, and Klayman (1998), Larrick (2004), and Kenyon (2014).

the opposite,' Norbert Schwartz and colleagues observe, 'the more they often convince themselves that their initial judgment was right on target' (2007: 128). In general, psychologists are fond of pointing out the inbuilt fallibility of debiasing techniques. Such techniques may be our best shot to debias, but that doesn't mean they're a good shot.

Remember, we wanted to know whether debiasing could let us reasonably accept (1), the premise that we are not biased, in spite of our awareness of the blind spot. Would-be debiasers face obstacles, as noted. Of course, some of us will still regard ourselves as having successfully debiased. But without instruction in debiasing and practice in implementing the techniques, aren't we just fooling ourselves? True, philosophers are trained to think carefully about arguments, but we are not trained how to debias for the blind spot. Again, rechecking our arguments and finding they still look good to us is not enough. At the very least, we should think it is unclear whether or not we've effectively debiased. The take-home message is that good reason for regarding ourselves as having debiased for the blind spot, and thus good reason for (1), is uncommon in the kind of disagreements we care about.

In the end, we might expect a person unknowingly subject to the bias blind spot to use the Self-Exculpating Strategy. But forewarned against the blind spot, we shouldn't deploy that strategy unless, again, we have good reason to think our self-judgment about bias is reliable.

In hopes of finding a better way to debunk our dissenters, let's turn to the

*Biasing-Factor-Attribution Strategy.* We have reasons to accept (1) that a factor *F* tends to bias judgments about proposition *p*, (2) that factor *F* holds for a disagreeing thinker's judgment about *p*, and (3) that we know of no 'enlightening' factor that holds for the thinker's judgment about *p* (i.e., a factor that tends to counteract or neutralize *F*'s influence). On the basis of (1) through (3), we infer that the Bias Premise is (probably) true.

How does this strategy look against the backdrop of the psychological picture? Here's a general worry to start with. One precondition for successful use of this strategy is that we lack a reason to accept that a biasing factor like *F* holds *for us*. Otherwise, the strategy will explode in our hands—it will debunk our dissenters and ourselves and fail to preserve our rational belief. But now we have reason to doubt that our perspective on ourselves is objective, given our inclination to cling to (unreliable) introspective evidence and disregard behavioral evidence concerning ourselves, for instance. Thus, we should sometimes suspect that the biases we attribute to others apply to our own judgment, too. Even if we set that point aside, both (1) and (2) are problematic for other reasons. I'll consider two.

First, our abstract theories of bias aren't always well-attuned to reality. We occasionally cynically overestimate bias in others (Kruger and Gilovich 1999). Let's suppose our theories say some factor *F* biases certain kinds of judgments. Then we should ask: on reflection, do we have good reason to think that our abstract theories of bias are right? Suppose we lack such reason. Then it would



be strange for us to accept (1): if we are unsure whether our abstract theories are correct, we should also be unsure about whether *F* really biases. This suggests that, once we reflect, reason to accept (1) amounts to reason to think that our abstract theories are reliable guides to bias, but we may often lack any reason to accept this.

Second, and shifting to (2), we appeal to behavioral evidence when determining whether some thinker is biased, but our uptake of that evidence may be influenced by unreliable evidence-gathering methods. Psychologists have noted that observers often place heavy weight on others' *characters* to explain their behavior in a particular situation, rather than thinking about how their behavior may have been shaped by the *situation* (Jones and Harris 1967). This has been called the 'fundamental attribution error' because it is arguably a central engine of social judgment. We explain how others act primarily by appeal to their characters, not to the situations they're in. When a bicyclist crashes, for example, we are apt to judge he's an unskilled or reckless rider. But that judgment is an error when the crash is explained instead by some part of the rider's situation. Maybe he's a good rider, but he was late for an appointment and knew he had to ride a bit dangerously around a corner to arrive on time. (Unsurprisingly, we readily regard situational factors as influencing our own behavior.) We should ask: on reflection, are the methods we use for collecting behavioral evidence concerning our disagreeing peers any good? If we lack reason to think those methods are good, it would be strange to accept (2): if we are unsure whether we can competently gather evidence relevant to factor *F* concerning a disagreeing thinker, we should also be unsure about whether *F* holds for that thinker's judgment. The idea here is that reason to accept (2) amounts to reason to think that our techniques for gathering behavioral evidence are reliable. But if we take the psychological research seriously, many of us often lack reason to think this.

The psychological picture raises doubts about whether we have good reason to accept (1) and (2), which sometimes calls into doubt our use of the Biasing-Factor-Attribution Strategy (cf. Christensen 2014: 160–61). At the same time, (3) is problematic, too. Recall that (3) is satisfied when we've reason to think that no enlightening factor—one that would counteract a biasing factor—holds for a disagreeing thinker. But this condition is too easy to meet. We could satisfy it, without fail, by remaining oblivious to the presence of potential enlightening factors. As a matter of fact, it's doubtful whether we are normally 'on the lookout' for enlightening factors operating in others who disagree with us. Given naïve realism—that is, the presumption that our views are objective—whenever others disagree with us, we tend to be on the lookout for biasing factors. The engine driving this may be confirmation bias. In general, we search for evidence to confirm our hypothesis that we are objective and not for evidence to disconfirm it; it's thus unsurprising that we would search for biasing factors, not for enlightening ones, because we expect that our dissenters are biased. As a result, we may take ourselves to accept (3) reasonably just because we've failed to search adequately for enlightening factors.

This point about (3) suggests a change to the Biasing-Factor-Attribution Strategy. Suppose that we reasonably accept premises (1), (2), and (3). Imagine further, as the psychological picture recommends, we reasonably believe we are

not always disposed to search for factors that may enlighten disagreeing thinkers, even when such factors are present. But this perhaps reveals that the strategy at issue doesn't set down conditions *sufficient* for us to have reason to accept the Bias Premise. If there's a gap, we may fill it with a further premise, namely, that we have reason to accept premise (4), that we have adequately searched for potential enlightening factors that hold for the disagreeing thinker's judgment about proposition *p*. Once again, the psychological picture should give us pause. Since it's doubtful we are in the habit of searching for factors that may enlighten our disputants, we have some reason to doubt whether we'd satisfy the condition in (4).

My suggestion is not that the Self-Exculpating Strategy and the Biasing-Factor-Attribution Strategy can't ever help us debunk testimony from those who disagree with us. The idea is that, in light of the psychological picture, we often have reason to doubt that these strategies do help. To make use of them in view of that picture, we must face up to the following questions. What is our reason to think that our self-judgments about biases are reliably formed? Why do we think our abstract theories of bias are any good? Why do we think our methods of collecting behavioral evidence about our dissenters are reliable? Why do we think we have adequately searched for factors that may enlighten disagreeing thinkers? To answer well, we must know how to make principled judgments about bias.

To be sure, sometimes we can answer those questions with relative ease and reasonably debunk dissenters. For instance, when we imagine Christensen's high-school music contest, it's natural to assume (1) that you've got good reasons to think your abstract theory of bias that implies Aksel's father is biased is in fact reliable, (2) that your gathering of behavioral evidence concerning Aksel's father's judgment has been reliable, and (3) that you've properly searched for factors that might enlighten him. Thus, you can launch the Biasing-Factor-Attribution Strategy and debunk Aksel's father. The same plausibly goes for the case of the climate scientists sponsored by the oil industry. But these are easy cases, and we are here concerned with less straightforward ones. In more difficult cases, we will on reflection have reason to doubt that these strategies are appropriate. While we may be tempted to see our intelligent and informed opponents as unable to hear the good sense we're talking, because of their social position or intellectual commitments, we now know that is too convenient. We may be inclined to treat ourselves as the unbiased side in disputes, to see ourselves as judge and jury, prosecution and defense, in our own case. But can't we do better? Ordinarily, we do not demand good reasons for our own judgments concerning biases even when we hold others to tougher standards. The psychological picture pushes us to resist the temptation to intellectual hypocrisy.<sup>12</sup>

<sup>12</sup> The arguments in this section are indebted to pioneering work by Kornblith (1999), Foley (2001: chap. 3), and Kelly (2008) on the epistemic implications of bias research. Christensen (2014) considers how some armchair judgments about bias ('unquantified common-sense psychology') can be used to react to disagreements, and I've benefited from his insightful discussion, too.

## 5. Conclusion

Let me close by highlighting two lessons. I began by noting how the debunking reasoning helps us build our intellectual houses. We may now begin to rip things apart. I said that to resist the downward push of testimony from those who disagree, we commonly resort to debunking reasoning. Though conciliationists and nonconciliationists disagree over the scope of rational disagreement—the range of cases where it is rational to keep one’s view while recognizing a peer’s dissent—both camps agree that sometimes controversial beliefs are rational for at least one peer because of debunking. And so both camps should acknowledge a problem: in light of the psychological picture, we often have powerful reasons to doubt that debunking delivers us from epistemic trouble. The upshot is that disagreement should lead us to reduce confidence in our controversial views more than we might have previously thought.

The first lesson is that without the safeguard of debunking reasoning, more intellectual modesty should be in store for us—we should hold some of our controversial beliefs with less confidence.

A second lesson is that we need better methods to make judgments about biases. We are invested in the practice of making such judgments, but our methods are makeshift, underwritten as they are by assumptions and inferences that have not been scrutinized nearly enough, let alone fully articulated. Improvement demands that we become more principled. I will briefly suggest what that may mean.

Many intellectual problems will not see conclusive resolution in our time. We will fail to reach the bottom. Sometimes issues remain unresolved because of their sheer complexity (see Ballantyne, forthcoming). Other times they remain unresolved because *we* are the ones trying to resolve them. We get in the way. Here is an image: a photographer is trying to take a photo of himself in the mirror without having the camera appear anywhere in the photo. It can’t be done. Similarly, in many disagreements, our judgments about biases in others and ourselves are fraught with difficulty that traces back to the presence of ourselves. Somebody’s thinking is either sensitive to evidence and reasons, or it is driven by his or her interests, expectations, or emotions. Who is biased? Me or you? Us or them? In many disagreements, we are not well-positioned to figure this out because our viewpoint is ours.

Perhaps that’s just how it goes: in good times and bad, we are stuck with ourselves. And yet we may hope for *impersonal* application of our methods. Not just methods that we will apply uniformly to everyone but ones that will bracket out the personal factors that bias application of those methods. Notice how the need for impersonal judgment in the shadow of potential bias has been met in one of our beloved academic institutions: recusal. College deans, department chairs, and journal editors may recuse themselves from decision making because of possible bias and thereby preserve justice and fairness. These practices respect an insight: if we can’t be trusted to apply methods properly, we should step back and let others do it. We insist on such practices because we want justice both to be done and to appear to be done. Likewise, in intellectual life, as we try to improve our webs of belief, we need a way to step back and to respect the fact that our views are often no less subject to the same biases we so readily attribute to others. But we

can't always be trusted to decide when to bow out, and so better methods might effectively exclude us at the right times.

How might this work? I am unsure. But as I conclude, let me offer a speculation. Psychologists have noted that people will occasionally recognize bias in themselves. Subjects have been observed to accept the idea that they are biased in their judgments 'broadly and abstractly construed' while at the same time disavowing bias in any recent judgments they've made (Ehrlinger, Gilovich, and Ross 2005). More important, subjects sometimes go a step further, confessing that some of their specific judgments are biased. For instance, people will admit to being biased in their assessments of their friends, and parents will admit they are biased toward their children. Some psychologists have proposed here that 'the motivation to be seen as unbiased is not as great—or is balanced by a countervailing motive to be a stand-up friend or a protective parent—so it is easier to admit to the possibility of bias' (Ehrlinger, Gilovich, and Ross 2005: 690).

As someone who often feels unbiased, but who knows, in quiet moments, that this feeling must be a hard-to-see-through cognitive illusion, I find good news here. The good news is that human motives and impulses may counteract the powerful tendency to see ourselves as unbiased. I take comfort in this and hope one day to see myself more as I really am: biased. What motive could help me? Again, I am unsure. Yet, suppose it becomes my central purpose as a thinker to simply consider how things are. Not to judge. Not to conclude. But to abstain from judgment and opinion. To try, as Ralph Waldo Emerson once put it, 'to keep the balance true'. To adopt this motive is to embrace thoroughgoing, non-dogmatic modesty in some controversies. Could the motive to consider how things are, just like my affection for a friend or a child, help me to more often recognize my biases by counteracting my tendency to resist the thought that I'm biased? Possibly so—if the motive is great enough. But perhaps this too is wishful thinking.

NATHAN BALLANTYNE  
 FORDHAM UNIVERSITY  
[n.ballantyne@gmail.com](mailto:n.ballantyne@gmail.com)

## References

- Arnauld, Antoine, and Pierre Nicole. ([1683] 1996) *Logic or the Art of Thinking*. Translated by Jill Vance Buroker. Cambridge, UK: Cambridge University Press.
- Ballantyne, Nathan. (2014) 'Knockdown Arguments'. *Erkenntnis*, 79, 525–43.
- Ballantyne, Nathan. (Forthcoming) 'The Significance of Unpossessed Evidence'. *The Philosophical Quarterly*. doi: 10.1093/pq/pqu096.
- Bergmann, Michael. (2009) 'Rational Disagreement after Full Disclosure'. *Episteme*, 6, 336–53.
- Christensen, David. (2007) 'Epistemology of Disagreement: The Good News'. *The Philosophical Review*, 116, 187–217.
- Christensen, David. (2010) 'Higher-order Evidence'. *Philosophy and Phenomenological Research*, 81, 185–215.
- Christensen, David. (2011) 'Disagreement, Question-Begging and Epistemic Self-Criticism'. *Philosophers' Imprint*, 11, 1–22.

- Christensen, David. (2014) 'Disagreement and Public Controversy'. In Jennifer Lackey (ed.), *Essays in Collective Epistemology* (New York: Oxford University Press), 143–63.
- Christensen, David, and Jennifer Lackey, eds. (2013) *The Epistemology of Disagreement: New Essays*. New York: Oxford University Press.
- Cross, K. Patricia. (1977) 'Not Can, But Will College Teaching be Improved?' *New Directions for Higher Education*, 17, 1–15.
- Ehrlinger, Joyce, Thomas Gilovich, and Lee Ross. (2005) 'Peering into the Bias Blind Spot: People's Assessments of Bias in Themselves and Others'. *Personality and Social Psychology Bulletin*, 31, 680–92.
- Elga, Adam. (Unpublished manuscript) 'Lucky to be Rational'.
- Feldman, R. (2005) 'Respecting the Evidence'. *Philosophical Perspectives*, 19, 95–120.
- Feldman, Richard, and Ted Warfield, eds. (2010) *Disagreement*. New York: Oxford University Press.
- Foley, Richard. (2001) *Intellectual Trust in Oneself and Others*. New York: Cambridge University Press.
- Frances, Bryan. (2013) 'Philosophical Renegades'. In David Christensen and Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays* (New York: Oxford University Press), 121–66.
- Fumerton, Richard. (2010) 'You Can't Trust a Philosopher'. In Richard Feldman and Ted Warfield (eds.), *Disagreement* (New York: Oxford University Press), 91–110.
- Haldane, J. B. S. (2002) *Possible Worlds*. New Brunswick, NJ: Transaction Publishers.
- Haldane, J. S., A. M. Kellas, and E. L. Kennaway. (1919) 'Experiments on Acclimatization to Reduced Atmospheric Pressure'. *Journal of Physiology*, 53, 181–206.
- Haldane, J. S. (1922) *Respiration*. New Haven, CT: Yale University Press.
- Heath, Chip, Richard Larrick, and Joshua Klayman. (1998) 'Cognitive Repairs: How Organizational Practices can Compensate for Individual Shortcomings'. *Research in Organization Behavior*, 20, 1–37.
- Jones, E. E., and V. A. Harris. (1967) 'The Attribution of Attitudes'. *Journal of Experimental Social Psychology*, 3, 1–24.
- Kahneman, Daniel. (2003) 'A Perspective on Judgment and Choice'. *American Psychologist*, 58, 697–720.
- Kelly, Thomas. (2008) 'Disagreement, Dogmatism, and Belief Polarization'. *Journal of Philosophy*, 105, 611–33.
- Kelly, Thomas, and Sarah McGrath. (Forthcoming) 'Are There Any Successful Philosophical Arguments?'. In John A. Keller (ed.), *Being, Freedom and Method: Themes from van Inwagen* (Oxford: Oxford University Press).
- Kenyon, Tim. (2014) 'False Polarization: Debiasing as Applied Social Epistemology'. *Synthese*, 191, 2529–47.
- King, Nathan. (2012) 'Disagreement: What's the Problem? Or A Good Peer is Hard to Find'. *Philosophy and Phenomenological Research*, 85, 249–72.
- Kornblith, Hilary. 1999. 'Distrusting Reason'. *Midwest Studies in Philosophy*, 23, 181–96.
- Kruger, Justin, and Thomas Gilovich. (1999) 'Naïve Cynicism' in Everyday Theories of Responsibility Assessment: On Biased Assessments of Bias'. *Journal of Personality and Social Psychology*, 76, 743–53.
- Larrick, Richard. (2004) 'Debiasing'. In Derek Koehler and Nigel Harvey (ed.), *Blackwell Handbook of Judgment and Decision Making* (Malden, MA: Wiley-Blackwell), 316–37.
- Lewis, David. (1983) *Philosophical Papers*. Vol. 1. New York: Oxford University Press.
- Nisbett, Richard, and Timothy Wilson. (1977) 'Telling More Than We Can Know: Verbal Reports on Mental Processes'. *Psychological Review*, 84, 231–59.
- Pronin, Emily. (2007) 'Perception and Misperception of Bias in Human Judgment'. *Trends in Cognitive Sciences*, 11, 37–43.
- Pronin, Emily, Daniel Y. Lin, and Lee Ross. (2002) 'The Bias Blind Spot: Perceptions of Bias in Self versus Others'. *Personality and Social Psychology Bulletin*, 28, 369–81.

- Pronin, Emily, and Matthew Kugler. (2007) 'Valuing Thoughts, Ignoring Behavior: The Introspection Illusion as a Source of the Bias Blind Spot'. *Journal of Experimental Social Psychology*, 43, 556–78.
- Putnam, Hilary. (1982) *Reason, Truth, and History*. New York: Cambridge University Press.
- Reeder, Glenn D., John B. Pryor, Michael J. A. Wohl, and Michael L. Griswell. (2005) 'On Attributing Negative Motives to Others Who Disagree With Our Opinions'. *Personality and Social Psychology Bulletin*, 31, 1498–1510.
- Ross, Lee, Joyce Ehrlinger, and Thomas Gilovich. (In press) 'The Bias Blindspot and its Implications'. In K. D. Elsbach and A. B. Kayes (eds.), *Contemporary Organizational Behavior in Action* (Upper Saddle River, NJ: Pearson Prentice Hall).
- Russell, Bertrand. ([1928] 2004) *Sceptical Essays*. London: Routledge.
- Schwartz, Norbert, Lawrence J. Sanna, Ian Skurnik, and Carolyn Yoon. (2007) 'Metacognitive Experiences and the Intricacies of Setting People Straight'. *Advances in Experimental Social Psychology*, 39, 127–61.
- Sorensen, Roy. (1988) *Blindspots*. Oxford: Clarendon Press.
- Taylor, Shelley, and Jonathon Brown. (1988) 'Illusion and Well-Being: A Social Psychological Perspective on Mental Health'. *Psychological Bulletin*, 103, 193–210.
- Wilson, Timothy, and Nancy Brekke. (1994) 'Mental Contamination and Mental Correction: Unwanted Influences on Judgments and Evaluations'. *Psychological Bulletin*, 116, 117–42.
- Wilson, Timothy, David Centerbar, and Nancy Brekke. (2002) 'Mental Contamination and the Debiasing Problem'. In Thomas Gilovich, David Griffin, and Daniel Kahneman (eds.), *Heuristics and Biases* (New York: Cambridge University Press), 185–200.
- Winters, Barbara. (1979) 'Believing at Will'. *Journal of Philosophy*, 76, 246–56.
- Van Inwagen, Peter. (2009) *Metaphysics*. 3rd ed. Boulder, CO: Westview Press.