CrossMark

Episteme (2022), **19**, 66–88
doi:10.1017/epi.2020.9

episteme

ARTICLE

An Interactionist Approach to Cognitive Debiasing

Steven Bland

Huron University College, London, Ontario, Canada
Email: sbland2@uwo.ca

(Received 17 September 2019; revised 21 January 2020; accepted 8 February 2020;
first published online 4 May 2020)

Abstract

This paper examines three programmatic responses to the problem of cognitive bias: virtue epistemology, epistemic paternalism, and epistemic collectivism. Each of these programmes focuses on a *single* level of epistemic analysis: virtue theorists on individuals, paternalists on environments, and collectivists on groups. I argue that this is a mistake in light of the fact that cognitive biases arise from *interactions* between these three domains. Consequently, epistemologists should spend less time defending these programmes, and more time *coordinating* them. This paper offers empirically based arguments for the interactionist approach, and contends that its adoption is an essential step for minimizing bias in empirical science.

Keywords: cognitive bias; interactionism; epistemic virtues; epistemic paternalism; epistemic collectivism

Mark Alfano and Joshua August Skorburg have recently argued for the use of an interactionist framework for bringing psychological research to bear on questions in ethics, epistemology, and the philosophy of mind (Alfano 2016; Skorburg 2017; Skorburg and Alfano 2019). The purpose of this paper is to argue that epistemic programmes dedicated to ameliorating the problem of cognitive bias should be developed and assessed within such an interactionist framework.

Interactionism is a view about the roles of personal traits and situational factors in giving rise to human behaviours. It is opposed to a view that I will call *isolationism*, according to which these two sorts of factors elicit behaviours *independently* of one another (person-situation independence). Kilhlstrom (2013: 794) argues that this supposition was responsible for the longstanding isolation of personality and social psychology from one another, as well as their longstanding debate about which type of factor is more influential: personal (the Doctrine of Traits) or situational (the Doctrine of Situationism). Interactionism is opposed not only to the Doctrines Traits and Situationism, but to the isolationist framework more generally. It is not the view that personal and situational factors play an equal role in generating behaviours, but that their effects are *co-dependent* and *non-linear*. Interactions between personal traits and situational features give rise to behaviours that cannot be adequately explained or predicted by the straightforward conjunction of theories regarding these two domains. Interactionism has thus given rise to more complex theories of social cognition.

© The Author(s), 2020. Published by Cambridge University Press

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

Cognitive bias has been a central focus of psychologists over the last five decades. The resulting heuristics and biases literature has become an important resource for epistemologists who aim to diagnose the sources of biased cognition and provide guidance to mitigate it. Virtue theorists argue that biases result from our over-reliance on intuitive forms of cognition. They claim that this shortcoming can be attenuated by cultivating responsibilist virtues, such as epistemic humility and self-vigilance. Epistemic paternalists insist that these strategies succumb to the very problems they are meant to solve: cognitive biases and motivated reasoning interfere with our efforts to debias our own thinking. They argue that a more effective approach is to identify and manipulate the situational factors that bias our cognition. Finally, collectivists contend that cognitive bias is a bigger problem for isolated agents than it is for epistemic groups, in large part because critical deliberation is more effective at reducing bias than character-based training or paternalistic constraints. Their opponents are quick to point out that cognitive biases are just as likely to be amplified in group settings as they are to be mitigated.

As was the case in psychology, the isolationist assumption of person-situation independence has led each of these programmes to focus on a *single* level of epistemic analysis: virtue theorists on individuals, paternalists on environments, and collectivists on groups. This is a mistake. Since our cognition is biased in many different ways, and for several different reasons, no single approach will suffice. But it would also be a mistake to pursue all three programmes independently; the fact that cognitive biases arise from *interactions* between individuals and their social and informational environments makes these ameliorative programmes interdependent. Moreover, they are not perfectly aligned: a strategy that is effective at one level can be ineffective, and even deleterious, at another. Thus, ameliorative programmes must be evaluated *holistically* in terms of how and when they satisfy their own ends, as well as the ends of the other two programmes. We must theorize not only about virtues, environments, and collectives, but about how these epistemic loci can be coordinated to yield synergistic effects. This paper offers empirically based arguments for the interactionist approach, and contends that its adoption is an essential step for minimizing bias in empirical science.

1. Virtue theories

Roberts and West (2015) and Samuelson and Church (2015) advocate similar programmes of cultivating responsibilist virtues to mitigate biased thinking. Both parties diagnose the source of cognitive biases within a dual-process framework that has enjoyed increasing support from psychologists in recent decades. In this framework, there are two types or systems of cognition: Type/System 1 and Type/System 2.[1] System 1 generates outputs automatically, effortlessly, unconsciously, and efficiently; it is at work, for example, when we understand simple sentences, or detect emotion in facial expressions. System 2 works consciously, voluntarily, effortfully, and slowly; its deliverances include solutions to complex calculations and evaluations of one's own conduct in social situations.

System 1 is highly efficient in part because of its use of *heuristics*, i.e., rules of thumb that generate judgements or decisions on the basis of little information. The availability heuristic, for example, gauges the frequency of an event by how easily its instances come to mind: I believe that there are more people under five feet tall than over seven feet tall because instances of the former are easier to think of than instances of the latter. On the other hand, System 2 requires more time and resources because it is sensitive to a

[1]Hereafter, I will refer to them as Systems 1 and 2, with the understanding that they are not actually physical or informational *systems*.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

broader range of information, which it processes by means of more complex reasoning. Any attempt to bring *statistical* information to bear on the question of the distribution of height in human beings will require the conscious operation of System 2. Because System 1 is so cost effective – its results are generally accurate and achieved with minimal effort in little time – it is our default mode of cognition. But its results can also be *systematically mistaken*. Reliance on the availability heuristic leads us to overestimate the frequency of salient, recent events, while underestimating the frequency of events that leave little impression. System 2 is more robustly reliable, and can scrutinize and override the deliverances of System 1 when it perceives the need to do so, but only by engaging significantly more cognitive resources. Our cognition thus involves a efficiency-accuracy trade-off, and biases result when we unduly favour efficiency over accuracy.[2] Samuelson and Church call this tendency the vice of *intellectual arrogance*:

> In a limited sense we might say that heuristics and biases exhibit the vice of "intellectual arrogance" because, they result, in part, from an inability to "decouple" from Type 1 representations and the thinker remains unable to leave his or her own perspective. (Samuelson and Church 2015: 1106)

In the process of cognitive decoupling, representations are decoupled from judgements (and decisions). This means that representations are not unreflectively endorsed, and become subject to rational reflection. Decoupling also involves entertaining alternative representations generated by hypothetical thinking. Construed in this way, it is easy to see how intellectual arrogance gives rise to cognitive biases: a failure to disengage from the representations of System 1 leaves us at the mercy of our systematically misleading intuitions.[3]

The good news is that intellectual arrogance can be overcome. Roberts and West, and Samuelson and Church conceptualize the process of doing so as involving the cultivation of virtuous character traits that moderate our more vicious tendencies: "We don't think natural heuristics (whether tricky or not) are virtues, but we do think character virtues can help correct our biased thinking" (Roberts and West 2015: 2562). And they claim that corrective virtues can be trained to a level of automaticity, such that they become traits of Systems 1 and 2 (Roberts and West 2015: 2562; Samuelson and Church 2015: 1108).

Roberts and West argue that biased thinking can be curbed by cultivating the virtues of self-vigilance and intellectual vitality:

> Self-vigilance is an ability intuitively to recognize cues, both mental and environmental, of our vulnerability to various kinds of epistemic mischance, to appreciate their significance, and to take defensive measures against them. Intellectual vitality, which is closely related to open-mindedness, is a kind of pre-emptive intellectual energy, enthusiasm, and attentive inquisitiveness. (Roberts and West 2015: 2574–5)

A self-vigilant person not only embraces fallibilism, but learns to appreciate the conditions in which they're susceptible to biased thinking. This appreciation necessarily

[2]This view has been opposed by the fast and frugal heuristics research programme. See, for example, Gigerenzer *et al.* (1999).

[3]Roberts and West identify four bias generating tendencies that arise when we fail to decouple: intellectual laziness, blinkered vision, associative coherence, and substitution. Since these all seem to be species of intellectual arrogance, Roberts and West's diagnosis of the source of cognitive biases is not substantially different from Samuelson and Church's.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

involves a recognition of the need to decouple from the representations of System 1 by engaging the costlier resources of System 2. Someone who is also intellectually vital will overcome their intellectual laziness in these circumstances by, for example, carefully considering the merits of views other than their own. These virtues are complementary in their tendency to mitigate cognitive biases: a vigilant person who lacks vitality will be too lazy to engage System 2 when they know they should, and a vital person who lacks self-vigilance is willing to forgo cognitive ease without knowing how and when to do so. To effectively serve their ameliorative function, then, self-vigilance and intellectual vitality must be developed in tandem.

Samuelson and Church argue that intellectual arrogance is best attenuated by cultivating *intellectual humility*, which they define as the tendency of "holding a belief as firmly as it is warranted" (2015: 1109). Since the process of properly calibrating our confidence requires that we decouple from the representations of System 1, their claim amounts to the truism that our failure to decouple can be overcome by decoupling when it is appropriate to do so. More specifically, they provide three strategies to facilitate intellectual humility. First, when we engage in rule-based thinking we are more likely to think analytically, and thus to decouple from the representations of System 1. When subjects are given criteria for attributing traits and abilities, their judgements about their own traits and abilities are less biased (in their favour) than subjects who are without criteria (Dunning *et al.* 2002). Second, when we hold ourselves and others accountable for the accuracy of our thinking, we increase our motivation to reason carefully and with an open mind. For example, subjects that know they're being judged on the accuracy of their assessments produce less biased judgements of their own abilities (Armor and Taylor 2002). Finally, our thinking tends to be less egocentric when we make a concerted effort to take on the perspectives of others. Subjects who were asked to entertain foreign viewpoints manifested less one-sided thinking than those who were unprompted to do so (Sedikides *et al.* 2007).

These ameliorative programmes are complementary: the interventions that Samuelson and Church recommend for cultivating intellectual humility – engaging in rule-based thinking, emphasizing accuracy and accountability, and taking on foreign perspectives – also seem like good ways of cultivating self-vigilance and intellectual vitality. Whatever virtues are nurtured by these means, Samuelson and Church agree with Roberts and West that they are of the *responsibilist* variety: "The presence or absence of these factors often hinges on motivation and the epistemic goals and values of thinking agents and thus favors the responsibilist or neo-Aristotelian account of virtue epistemology" (Samuelson and Church 2015: 1103). We can hold individuals responsible for their intellectual arrogance because it consists in a set of habits that can be changed with sufficient effort, resources, and motivation.

Both of these positions are problematic insofar as they neglect the role that *reliabilist* virtues must play in treating the problem of cognitive bias. The fact that reliabilist virtues must play such a role is made clear in the work of Keith Stanovich, who distinguishes two fundamental sources of biased thinking: cognitive miserliness and mindware problems (Stanovich 2011). Cognitive miserliness is intellectual arrogance under a different name, i.e., the failure to properly decouple from the representations of System 1. It can be mitigated by developing thinking dispositions that are constitutive of what Stanovich calls *fluid rationality*, but are also recognizable as responsibilist virtues, e.g.: resistance to miserly information processing (self-vigilance); actively open-minded thinking (intellectual humility); and the tendency to seek information (intellectual vitality). 'Mindware' is Stanovich's term for the cognitive processes that belong to System 2; these are the processes that can correct the deliverances of System 1 when they're biased, albeit at considerably more cognitive expense, e.g.: logical inference, probabilistic

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

reasoning, economic thinking, etc. When we lack the knowledge to properly engage these cognitive processes, we suffer from what Stanovich calls *mindware gap*. Missing mindware is responsible for many of the errors we make when reasoning about statistical phenomena: gambler's fallacy, non-regressive predictions, base rate neglect, etc. On the other hand, certain types of mindware are sources of systematic errors. Stanovich calls this *contaminated mindware*, which includes: superstitious thinking, belief in the superiority of intuition, and an overestimation of introspection. He distinguishes good mindware from contaminated mindware in terms of its tendency to produce accurate beliefs: "A different type of mindware problem arises because not all mindware is helpful – either to attaining our goals (instrumental rationality) *or to having accurate beliefs* (epistemic rationality)" (Stanovich 2011: 193, emphasis added). Mindware gap and contamination, then, are types of reliabilist vices. They must be redressed by developing *crystallized rationality*, which consists in the competences and knowledge required to reliably form and maintain accurate beliefs.

Stanovich's research thus favours a hybrid ameliorative programme that emphasizes *both* reliabilist *and* responsibilist virtues. Sustained decoupling is only as good as the mindware that System 2 uses throughout the process: to avoid conjunction errors, for example, we must not only avoid defaulting to System 1, but also possess the probabilistic knowledge required to identify them as errors, and correct them accordingly. But these virtues cannot be realized in the absence of dispositions to use such cognitive resources when they're needed. Crystallized rationality without fluid rationality is impotent; fluid rationality without crystallized rationality is blind.

Cognitive miserliness and mindware problems are *personal* failings that involve intellectual arrogance, ignorance, and incompetence. It makes sense to address these failings by cultivating personal virtues. If cognitive biases result from some systematic failures in the way we think, then we should change the way we think. Naturally, this is difficult, and sometimes impossible, for individuals to do, since we often don't know how to change our thinking, or even that there's anything wrong with our thinking in the first place. For this reason, we must be trained to think differently by those who know how to do so. This is how we acquire the mindware that's needed to overcome ignorance and incompetence: by learning about logic, probability, statistics, economics, scientific methodology, and the like, from those who are more knowledgeable and expert than we are. We must also learn to recognize how and when to use reliable mindware in situations where our intuitive judgements are likely to be contaminated by bias:

> If we are sensitive to situations in which a given bias is likely to victimize us, we will become intuitively suspicious of ourselves when in those situations, and so can take whatever measures of caution are appropriate to avoid the trap, at least much of the time. (Roberts and West 2015: 2563)

Roberts and West suppose that this self-vigilance can also be cultivated through sustained training: "Our proposal depends crucially on education" (2015: 2562). For example, students of critical thinking might be taught that they often fall prey to confirmation bias when thinking about topics on which they have existing views. They can be trained to counter this bias by using the *consider-the-opposite* strategy, i.e., considering some of the reasons why their judgements might be mistaken. An important part of this learning process will include exercises in which students deploy these metacognitive skills across a variety of concrete cases. With enough practice, they will learn to do so automatically, without having to be reminded and guided by an instructor.

This virtue theoretic programme is clearly an isolationist approach to the problem of cognitive bias. Indeed, this is one of its potential advantages: if individuals can cultivate

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

stable capacities and traits that attenuate biases across a wide variety of situations, then there is no need to rely on environmental and/or social interventions. Critics deny that this is the case, however, for reasons that will be examined in the following section.

2. Problems for virtue theories

The prospects of education-based virtue theoretic programmes seem dim for a straightforward reason: they require biased minds to carry out the process of debiasing. Research suggests that even well-motivated agents have significant difficulties accomplishing this feat. It also suggests that when we are motivated by factors other than accuracy, we often deploy our cognitive resources to *reinforce* our biases. These two findings will be considered in turn.

Since the cognitive processes that give rise to biases are largely unconscious and involuntary, we have no first-person access to their operation or effects.[4] Yet we operate under the assumption that we do have introspective access to the vast majority of our mental processes. So, when we look inwardly for signs of bias in our own thinking, and find none, we conclude that our thinking is unbiased. This is called *the objectivity illusion*. This impression is reinforced by the fact that System 1 processing is often accompanied by a feeling of cognitive ease, which we tend to interpret as a sign of its reliability (Kahneman 2011: 62). And having formed this impression, we preferentially seek and accept evidence in its favour. Ignorance and intellectual arrogance, then, are instances of what Cassam (2015) calls *stealthy vices*: vices that block their own detection. Missing or contaminated mindware often obscures the need for better mindware, and a failure to decouple when necessary often obscures the fact that it's necessary to decouple. These facts conspire to produce a *bias blindspot* in each one of us: the tendency to believe that we are immune to the biases that influence the judgements of others (Pronin *et al.* 2002, 2004; Ehrlinger *et al.* 2005). Consequently, while teaching people about cognitive bias may help them to identify biased thinking *in others*, it is unlikely to help them recognize those situations in which *their own* thinking is biased. Moreover, there is some evidence that by entertaining the possibility that our judgements are biased, we can actually make our biases *worse* (Sanna *et al.* 2002). In these situations, we are often subject to *phantom debiasing*: thinking that we've decontaminated our judgements, while only increasing our confidence in them (Frantz 2006). Finally, even when we do engage in debiasing, there are serious obstacles to effective decontamination. Properly calibrating our response to bias requires that we know its full effect on our judgements, and here we must rely once again on theorizing rather than introspection. This is yet another place for bias to creep into the debiasing process, leading to the common occurrences of insufficient correction, unnecessary correction, and overcorrection (Wilson *et al.* 2002: 191).[5]

This literature suggests the sobering conclusion that our metacognitive reasoning is least reliable when our cognition is least reliable: we may get things right most of the time, and know that we get things right most of the time, but when we get things wrong, we're less likely to know it, and less likely still to properly correct it. It looks, then, like intellectual humility, vitality, and self-vigilance are not the sorts of virtues that can be readily cultivated through training and habituation.

[4]See Nisbett and Wilson (1977), Wilson (2002), Wilson *et al.* (2002), Pronin *et al.* (2004), Ehrlinger *et al.* (2005) and Pronin (2007).

[5]While some debasing strategies have been used successfully by experimental subjects, critics argue that they are unlikely to be successfully deployed in less artificial settings (Trout 2005: 419–20; Ahlstrom-Vij 2013: 28; Kenyon and Beaulac 2014: 347).

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

The trouble doesn't end there for virtue epistemologies. Let's suppose, contrary to what's been said thus far, that the corrective virtues could be developed through prolonged training. Would such training render agents less susceptible to cognitive biases? While this may be the case for epistemically motivated agents, there is reason to believe that such training could have the *opposite effect* for agents with non-epistemic motivations (Kunda 1990). Dan Kahan (2013, 2017) has recently made this case with respect to ideologically motivated reasoning. Beliefs about such controversial topics as gun control and climate change are symbols of allegiance to our socio-political tribes. We don't often abandon these beliefs when presented with challenging empirical evidence; instead, we find reasons to dismiss the evidence or new reasons to maintain our convictions. The case of climate change scepticism makes this abundantly clear. Yet Kahan insists that this reaction does not constitute an irrational failing arising from an overreliance on System 1, as others have argued. Rather, it is an *expressively rational* strategy of maximizing expected utility. When someone correctly revises their identity-congruent beliefs, they become one more voice on the right side of an important socio-political issue, but this single conversion will make no difference at the level of policy. On the other hand, there is a good chance that it will end up alienating them from their affinity group. Consequently, members of opposing groups have every reason to toe the line on the issues that divide them, regardless of the state of relevant evidence, because being wrong is not as hazardous as being ostracized.

We're also motivated to think of ourselves as rational people. Therefore, it is not enough to believe what our confederates believe; we must also have *reasons* for our identity-congruent beliefs (Kunda 1990: 483). Our expressively rational aim, then, is to dogmatically hold on to these beliefs without looking, to ourselves or others, like we are doing so. This is no simple feat, which suggests that individuals with more cognitive resources are more likely to succeed at managing it than individuals with fewer resources:

> But *if* individuals *are* adept at using more effortful, System 2 modes of information processing, then they ought to be even *better* at fitting their beliefs to their group identities. Their capacity to make sense of more complex forms of evidence (including quantitative data) will supply them with a special resource that they can use to fight off counterarguments or to identify what stance to take on technical issues more remote from ones that that [sic] figure in the most familiar and accessible public discussions … More importantly still, it will make them more likely to *understand* the significance of competing claims, and related forms of evidence, for the status of their group, and thus be more likely to experience unconscious motivations to form identity-congruent assessments of them (Kahan 2013: 409–10).

And this is precisely what Kahan has found: subjects who score *higher* in numeracy and cognitive reflection are *more likely* to interpret evidence as favouring their political outlooks, and therefore tend to have *more polarized* views about politically divisive topics (Kahan 2013; Kahan *et al.* 2017). In this domain, the virtues that typically facilitate the attenuation of cognitive biases serve an antithetical purpose:

> The reason that citizens remain divided over risks in the face of compelling and widely accessible scientific evidence, this account suggests, is not that that [sic] they are insufficiently rational; it is that they are *too* rational in extracting from information on these issues the evidence that matters most for them in their everyday lives. (Kahan *et al.* 2017: 78)

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

Thus, attempts to make people less biased in the face of policy-relevant science by training them to be more numerate and reflective are unlikely to succeed. Indeed, this strategy may have the *opposite* effect. More generally, intellectual humility, vitality, and self-vigilance facilitate the correction of biased beliefs only when we are motivated primarily by a desire for accuracy; otherwise, they can *inhibit* this process.

All of this sounds like bad news for *any* ameliorative programme that hopes to mitigate the negative effects of biased cognition. A growing number of epistemologists insist that this isn't the case; rather, they argue, this is bad news only for traditional programmes that focus too narrowly on personal factors, at the expense of giving due consideration to situational ones. Epistemic paternalists are especially sensitive to the latter, and claim to offer ameliorative strategies that are unencumbered by the problems surveyed in this section.

3. Epistemic paternalism

If I'm struggling to see in a dark room, I don't try to improve my vision; I turn on the light. My visual perception is a reliable source of information, but only when operating in sufficiently well-lit environments. More generally, our epistemic successes are a function of both our cognitive processes and the environments in which they operate. To improve our chances of success, then, we can seek to make changes to our minds, or the relevant features in our environments. J.D. Trout labels these inside and outside strategies, respectively:

> An *inside strategy* is a voluntary reasoning process designed to improve the accuracy of judgment by creating a fertile corrective environment *in the mind.* (Trout 2005: 418)
> An *outside strategy* identifies features *of the environment* whose presence can be manipulated to produce the most accurate or desirable available outcome. (Trout 2005: 420)

Trout and others claim that inside debiasing strategies are generally fruitless, but outside strategies that seek to scaffold cognition are considerably more promising. They offer what amounts to a *situationist* critique of agent-based attempts to mitigate the problem of cognitive bias. On this view, traditional virtue-theoretic programmes are akin to attempts to improve our vision in the dark: their prospects are dim, and there's a much better way of addressing the problem.

Outside strategies are typically designed and implemented by experts and/or officials with the intention of improving peoples' cognition, sometimes without their knowledge or consultation. These are cases of *epistemic paternalism*. Some paternalistic measures seek to minimize biased cognition by imposing external constraints on peoples' access, collection, and evaluation of information (Ahlstrom-Vij 2013). One such constraint is exemplified in the US *Federal Rules of Evidence* which require judges to withhold information that may unfairly bias the jury's views of the defendant. Other paternalistic interventions seek to *harness* biased cognition to yield more accurate judgements (Trout 2005). Framing effects can be particularly powerful in this regard. Consider the following two problems (Trout 2005: 423–4):

> *Probability Format*
> For women at age 40 who participate in routine screening, the probability of breast cancer is 1%. If a woman has breast cancer, the probability is 80% that she will get a positive mammography. If a woman does not have breast cancer, the probability

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

> is 10% that she will get a positive mammography. Suppose a woman in this age group had a positive mammography in a routine screening. What is the probability that she actually has breast cancer? __%
>
> *Frequency Format*
> About 10 out of every 1000 women in this group (age 40–50 who participate in routine screening) have breast cancer. About 8 of every 10 women with breast cancer will get a positive mammography. Out of the 990 women without breast cancer, 99 will also get a positive mammography. Now you are introduced to a new representative sample of women at ages 40–50 with no symptoms who got a positive mammography in routine screening. How many of these women actually have breast cancer? __ out of __.

They are, of course, the same problem. But Gigerenzer (1996) found that only 16% of statistically untrained subjects answer probability questions correctly, while 46% answer questions about relative frequencies correctly. That's nearly a threefold improvement without any statistical instruction. Furthermore, there is evidence that even individuals with training in probabilistic reasoning, such as faculty and staff at Harvard's Medical School, have difficulty applying Bayes' rule when reasoning about conditional probabilities (Casscells *et al.* 1978). It's clear why people have an easier time with frequency formats than with probability formats: the calculations are easier, and they make it more obvious how to use information about base rates. The use of frequency formats has also been found to drastically reduce the incidence of overconfidence and conjunction errors (Gigerenzer 1991; Gigerenzer and Hoffrage 1995; Hertwig and Gigerenzer 1999). Implementing this strategy doesn't involve withholding any information, or controlling how it is evaluated; it simply involves *framing* the information in a way that is generally conducive to more accurate, less biased cognitive processing.

Recall that Kahan denies that peoples' thinking about politically charged topics can be effectively debiased by means of training that inculcates corrective virtues; in fact, it seems that this training tends to make peoples' thinking about these topics *more* biased, not less. So what can be done to attenuate the motivated reasoning that contaminates our thinking on these topics? Kahan *et al.* insist that, "Identifying strategies for protecting the science communication *environment* from antagonistic cultural meanings – and for decontaminating it when such protective measures fail – is the most critical contribution that decision science can make to the practice of democratic government" (Kahan *et al.* 2017: 80, emphasis added). The findings of policy relevant science cannot be withheld from the public, nor can the public be forced to reason about them rationally. Instead, the communicators of this science might try framing what are bound to be highly contested facts in less polarizing ways.

To be clear, paternalists do not deny that epistemic vices play a central role in biased cognition. But they insist that such personal factors are only part of the story; a fuller explanation must also acknowledge the role of hostile environments. And because it is easier to improve our informational environments than it is to remedy the stealthy vices that give rise to cognitive biases, we should favour outside strategies of epistemic nudging over inside strategies of personal training.

Undoubtedly, outside strategies sometimes succeed where inside strategies fail. Nevertheless, there are good reasons not to wholeheartedly endorse the paternalist programme. The first is that it rests on an overly pessimistic view about the prospects of inside strategies. More importantly, its isolationist approach of focussing on situational interventions rather than personal interventions obscures the extent to which situational and personal factors are co-dependent; by optimizing their interactions, we

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

can achieve much more than we can by inside and outside strategies alone. Both of these points are elaborated in the following section.

4. Problems for paternalism

The arguments for pessimism examined in §2 do not apply equally to all inside strategies. Recall that Stanovich distinguishes between two types of cognitive shortcomings: cognitive miserliness and mindware problems. While the former results from deficits of responsibilist virtues, the latter are attributable to a lack of reliabilist virtues. It is entirely possible that inside strategies can effectively cultivate one type of virtue, but not the other. The arguments in §2 suggest that the responsibilist virtues touted by Roberts and West (2015) and Samuelson and Church (2015) cannot be developed through training because lessons get applied in biased ways at the metacognitive level. But this does not imply that the mindware gap cannot be closed by well designed training. This gap partially accounts for many of the biases we manifest when reasoning about statistical matters: the law of small numbers, non-regressive predictions, probability matching, etc. Can these errors not be reduced through statistical training? It seems they can. In both longitudinal and cross-sectional studies, Lehman *et al.* (1988) found that third-year graduate students in psychology and medicine showed marked improvements in statistical reasoning over students in their first year: the frequency and quality of their statistical answers to scientific and everyday questions increased dramatically.[6] They found no significant improvement in the performance of law and chemistry students from first to third year; this discrepancy is readily explained by the fact that students of psychology and medicine receive statistical training, while students of law and chemistry generally do not. Lehman and Nisbett (1990) discovered a similar trend among undergraduate students. Their longitudinal study found that the statistical and methodological reasoning of students in psychology and the social sciences improved much more from their first- to fourth-years than the reasoning of students enrolled in the humanities and natural sciences. Finally, Fong *et al.* (1986) found that subjects who were trained to use the law of large numbers, by means of either formal instruction or the use of examples, were more likely to do so successfully than subjects in the control condition. Furthermore, subjects proved adept at applying the law outside of the domain of their instruction. These gains remained, albeit diminished, when the subjects were tested again two weeks after their instruction. Thus, it appears that reliabilist virtues can be learned.

Further grounds for optimism come from Philip Tetlock's work on forecasting in socio-political domains. His early work revealed that experts are shockingly inept at making accurate forecasts, in part because of their unchecked biases. And he was sceptical that debiasing training could do much to improve their performance (Tetlock 2005: 189). More recently, he has studied the performance of non-experts in the *Good Judgement Project*, a component of a forecasting tournament funded by the Intelligence Advanced Research Activity Projects (IARPA). Tetlock was particularly interested in the participants that outperformed the vast majority of their peers, a group that he calls 'superforecasters'. He found that superforecasting is a stable skill that covaries with a number of cognitive traits, including such responsibilist virtues

[6]The scores of psychology students rose by nearly 70% in both the cross-sectional and longitudinal studies; the scores of medical students rose by roughly 25% in both studies. The statistical reasoning portion of the test asked questions that were best answered by applying the law of large numbers and the regression principle across a broad domain of problems, many of which fell outside of the domains in which students were used to applying the principles.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

as: epistemic humility; caution; active open-mindedness; reflection; grit; fallibilism; and a growth mindset (Tetlock and Gardner 2015: 191–2). He is now more optimistic about the prospects of nurturing these virtues: "These habits of thought can be learned and cultivated by any intelligent, thoughtful, determined person" (Tetlock and Gardner 2015: 18). This turnabout is largely due to his experience developing and administering a training programme meant to improve participants' forecasting accuracy. The programme consists in a one-hour tutorial that was found to increase trainees' accuracy by 10% the following year. Furthermore, Tetlock surmises that superforecasters have learned to *automatically* apply the techniques covered in the tutorial: "My sense is that some superforecasters are so well practiced in System 2 corrections – such as stepping back to take the outside view – that these techniques have become habitual. In effect, they are now part of their System 1" (Tetlock and Gardner 2015: 236). If Tetlock's speculation is accurate, then responsibilist virtues may not only be learned and practiced, but intuitively realized in just the way that virtue epistemologists claim.

Paternalists will reasonably insist that we not ignore the role that situational factors are playing in the cultivation of these virtues. Though instruction improved forecasting performance, it did so in the context of a tournament that provides participants with clear, timely feedback about the accuracy of their judgements. Without this feedback, it could be argued, there would have been little or no improvement. In other words, responsibilist virtues facilitate forecasting success, but they're also facilitated by participation in forecasting tournaments. This supposition is consistent with the fact that superforecasters improved year over year *without* the benefit of instruction, and often improved *more* than ordinary forecasters (Tetlock and Gardner 2015: 102–3). Perhaps the best way to change the environment inside peoples' minds – making them less confident, more open-minded, etc. – is to improve the environments outside their minds. We might call these *outside-in* strategies.

Samuelson and Church recognize the possibility of using such strategies to induce intellectual humility:

> On the one hand, those low in the trait could cultivate the habits of intellectual humility through practice (while those high in the trait could grow in expertise through the same means). On the other, creating the right pressures and circumstances (accountability, etc.) could induce intellectual humility (or attenuate intellectual arrogance and intellectual diffidence). Like the state-trait debate in personality and social psychology, intellectual humility is neither simply state or trait but the result of the complex interaction of person and environment that is a part of all personality development. (Samuelson and Church 2015: 1108)

Of course, to say that persons and environments *interact* is to imply that causal influences run in *both* directions: persons also have some measure of control over their environments. This is a point that gets overlooked by situationists and paternalists, and it leads to my final objection. Like all forms of paternalism, epistemic paternalism runs the danger of minimizing individual agency: the more we rely on paternalistic interventions to mitigate cognitive biases, the less individuals will be able to improve their own cognitive processing. Yet not every instance of scaffolded cognition need be the result of a paternalistic outside strategy: agents can scaffold *their own* cognition in virtuous ways. If I am worried about biases influencing my grading, I can anonymize assignments and use grading rubrics, rather than trying to change my attitudes. If I want to manifest the virtue of open-mindedness, I can use the consider-the-opposite strategy, or I can seek out a conversation with critical interlocutors. These are *inside-out* strategies. And, as Kenyon and Beaulac rightly emphasize, they can and should be an

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

important part of any education that helps individuals to address their own biases: "critical thinking education should include extensive practical guidance on how to structure and engage with one's environment to promote good reasoning" (Kenyon and Beaulac 2014: 360). Indeed, this sort of guidance is essential to the virtue-based education envisioned by Roberts and West.

At the core of many virtue-theoretic responses to the problem of cognitive bias is a falsely optimistic view about the cultivation of epistemic virtues. Their proponents fail to recognize that biases operating at the metacognitive level make many of their inside strategies intractable, especially those that target responsibilist virtues. Paternalists, on the other hand, are too quick to overlook the fact that agents can develop intellectual virtues by designing and exploiting environmental interventions, rather than relying exclusively on their own cognitive resources. In short, neither of these isolationist positions gives due consideration to the mind-world interactions that give rise to cognition.

Collectivists argue that these theories pay insufficient attention to the role that mind-mind interactions play in our epistemic practices. They insist that we look to these interactions for the most promising debiasing strategies.

5. Collectivist epistemologies

Studies in the heuristics and biases tradition tend to focus on individual performance on abstract reasoning tasks. Yet much of our reasoning is done in communication with others, often in contexts where we are attempting to convince them of the truth of our beliefs or the wisdom of our decisions. Mercier and Sperber (2011, 2017) have recently argued that reason evolved in human beings to facilitate effective communication, rather than to optimize the judgements and decisions of individuals. One impediment to communication is that we are less likely to accept information that is surprising or in violation of our existing beliefs than information that is consistent with our world view. On the one hand, this is a good thing: we know (or believe) that we are a trustworthy source of information, but we often don't know this about others. Thus, privileging our own convictions is an effective means of avoiding being misled or deceived. On the other hand, if we accepted information from others only when it is consistent with what we already believe, testimony would seldom expand or improve our view of the world. Our capacity to produce and evaluate reasons evolved to enable us to avoid these excesses of testimonial scepticism and credulousness: we can argue for our beliefs when others are disinclined to accept them. In this way, argumentation increases the volume and fidelity of the information that gets communicated. The extent to which reason serves this *social* function cannot be assessed by evaluating the performance of isolated individuals on abstract reasoning tasks. Consequently, Mercier and Sperber argue, the effectiveness of reason has been mismeasured by the heuristics and biases tradition: "Human reasoning is not a profoundly flawed general mechanism; it is a remarkably efficient specialized device adapted to a certain type of social and cognitive interaction at which it excels" (Mercier and Sperber 2011: 72).

Defenders of the heuristics and biases tradition could argue, as some of Mercier and Sperber's respondents do, that reason serves *both* social interactionist *and* individual intellectualist ends, in which case experimental demonstrations of systematic irrationality at the level of individuals are relevant to assessments of its general efficacy (Evans 2011; Frankish 2011). This view, to which I am sympathetic, suggests the more moderate conclusion that reason is profoundly flawed in some respects and remarkably efficient in others. More importantly for our purposes, it suggests that reasoning can be *improved* by leveraging the social contexts in which it excels. Both of these conclusions are inherent in Mercier and Sperber's analysis of confirmation bias.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

Confirmation bias has deleterious effects on the ability of isolated subjects to solve abstract reasoning tasks. For example, when attempting to solve the Wason selection task or evaluate categorical syllogisms, subjects are more likely to find reasons in favour of their intuitive judgements than against them. Since peoples' intuitions about these cases are generally incorrect, the majority of subjects end up giving wrong answers. These facts are difficult to square with the intellectualist view that reason evolved to optimize our judgements and decisions. It is a perfectly natural consequence of the argumentative theory, however: if reason evolved to facilitate rational persuasion, then it should be biased in favour of the views it argues for. But if argumentation evolved as a social mechanism for vetting information, then confirmation bias should not interfere with this process. And indeed it doesn't, argue Mercier and Sperber. In fact, they claim that confirmation bias at the level of individuals produces an effective division of cognitive labour at the level of deliberating collectives:

> When a group has to solve a problem, it is much more efficient if each individual looks mostly for arguments supporting a given solution. They can then present these arguments to the group, to be tested by the other members … This joint dialogic approach is much more efficient than one where each individual on his or her own has to examine all possible solutions carefully. The advantages of the confirmation bias are even more obvious given that each participant in a discussion is often in a better position to look for arguments in favor of his or her favored solution (situations of asymmetrical information). (Mercier and Sperber 2011: 65)

This efficient division of cognitive labour depends on a collective's having certain features: viewpoint diversity; a common interest in the truth; at least one member with the ability to discern the truth; other members who are capable of being persuaded by good arguments. When these conditions are met, confirmation bias largely disappears at the level of collectives. This is why, for example, deliberating groups are several times more likely to correctly answer the Wason selection task than individuals (Moshman and Geil 1998; Augustinova 2008). Every group member is likely to argue in favour of their own solution (confirmation bias), but if members favour different solutions, and everyone is sensitive to good arguments, then groups will likely find the correct solution as long as one of their members does so. This suggests that neither the responsibilist vices that give rise to biased cognition nor the responsibilist virtues that attenuate them are summative in deliberating collectives: a collection of miserly cognizers can make up vigilant and vital groups.[7] Given the difficulty of cultivating these virtues at the level of individuals, it would seem that our efforts to mitigate biased cognition would be better spent cultivating productive group deliberations:

> So group discussions provide a much more efficient way of holding the confirmation bias in check. By contrast, the teaching of critical thinking skills, which is supposed to help us overcome the bias on a purely individual basis, does not seem to yield very good results (Ritchhart and Perkins 2005; Willingham 2008). (Mercier and Sperber 2011: 65)

Individuals can be taught to implement the consider-the-opposite strategy to mitigate the effects of confirmation bias (as well as other biases) on their cognition, but the

[7] As Mercier and Sperber emphasize, cognitive labour gets efficiently divided when group members are miserly in the *production* of arguments, but not in the *evaluation* of arguments. Their survey of the empirical literature on this topic suggests that subjects are naturally disposed to use their cognitive resources in these ways (Mercier and Sperber 2017: Ch. 12).

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

prospects of their doing so successfully appear dim in light of the biases that are likely to operate in their application of the strategy (Kenyon and Beualac 2014: 347). On the other hand, the biases of individuals can lead to an efficient and effective division of cognitive labour when the strategy is naturally implemented by deliberating groups.

These conclusions about group cognition are reinforced by Tetlock's finding that teams of forecasters outperformed individuals by 23% (Tetlock and Gardner 2015: 201). Moreover, he found that grouping superforecasters yields even more remarkable improvements in predictive accuracy:

> On average, when a forecaster did well enough in year 1 to become a superforecaster, and was put on a superforecaster team in year 2, that person became 50% more accurate. An analysis in year 3 got the same result. Given that these were collections of strangers tenuously connected in cyberspace, we found that result startling. (Tetlock and Gardner 2015: 205)

Furthermore, Tetlock reports that virtues related to superforecasting, such as active open-mindedness, are *emergent features* of epistemically motivated groups: forecasting teams can be open-minded, even when their constituent members are not (Tetlock and Gardner 2015: 207–8). If Mercier and Sperber are correct, then teams might be *more likely* to manifest active open-mindedness when their constituent members do not. At any rate, this is yet more evidence that favours the adoption of collectivist strategies over individualistic ones.

This collectivist ideal of improving reasoning by means of public deliberation is at the heart of positive visions of democratic policymaking (Habermas 1990; Mill 2003 [1859]) and scientific objectivity (Longino 1990; Popper 2002 [1996]). However, paternalists are less sanguine about the prospects of collectivist programmes; they point to a vast empirical literature that shows that collective deliberation can also introduce biases, and amplify existing ones (Sunstein 2003, 2009). Discussions on social media often bear out this concern (Sullivan *et al.* Forthcoming). In order to be successful, then, collectivist programmes must be able to identify the features of deliberating groups that tend to ameliorate rather than exaggerate cognitive biases. These features include personal and situational variables that have been largely ignored by the isolationist approach of most collectivists.

6. Problems for collectivists

Group deliberation outperforms individual cognition when it effectively aggregates information and entertains a variety of positions. In order for this to happen, participants must have non-overlapping knowledge, convictions, and justifications, and be willing to share them with the group. Even when there is the requisite diversity, a number of dynamics can interfere with its expression in deliberative contexts. Foremost among them is the phenomenon of *cascades*, which happen when individuals ignore their own knowledge in favour of the views expressed by others (Lorenz *et al.* 2011). This can happen because these individuals fear the social repercussions of voicing a dissenting view (reputational cascade), or because they treat the testimony of others as sufficient reason to silence themselves (informational cascade). In addition, groups often overweight the information that their members hold in common, which is known as the *common knowledge effect* (Gigone and Hastie 1993). When common knowledge favours a position that is contravened by information that is suppressed by cascades, groups suffer from *groupthink*: the tendency to arrive at an insufficiently reasoned consensus. It also pushes participants towards more radical versions of the views they held

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

prior to deliberation, resulting in *group polarization* (Baron *et al.* 1996). Furthermore, there is reason to believe that reasoning that is motivated to safeguard identity-congruent beliefs increases in collectivist contexts, since the social costs of abandoning such beliefs will be more salient in those situations.

Thus, having individuals deliberate with one another is no straightforward antidote to cognitive bias. When deliberations are short on diversity and/or criticism, groups can be just as biased as individuals, and sometimes more so. Progressive deliberation requires measures that hold groupthink and group polarization in check.

Deliberative polling has proven very successful in this regard. Deliberative polling is a process of public consultation wherein participants render their views on a specific political issue after having engaged in a series of moderated discussions with peers and experts. It begins by administering a questionnaire to a random, representative sample of the relevant population. From this sample, a smaller representative sample is randomly selected and given a package of topic-relevant information whose contents are vetted by a committee of stakeholders. Once informed, participants are randomly assigned to small groups, where they engage in moderated discussions among themselves, and plenary sessions in which they jointly question experts and policymakers. At the conclusion of the process, another questionnaire is administered to each participant, to capture their considered judgements on the topic.

By comparing participants' answers on the pre-deliberation and post-deliberation questionnaires, advocates have been able to show that deliberative polls (DPs) do facilitate critical discussions while attenuating many of the biases that afflict group cognition: participants often change their views, but not routinely in conformity with one another, or towards a more extreme position (Fishkin 2009: 131–2). These results are particularly surprising in light of the polarizing topics that get discussed, such as: the treatment of Aboriginals in Australia (2001), the treatment of the Roma in Bulgaria (2007), and educational reforms in Northern Ireland (2007). How do DPs manage this feat? By creating *environments* that are designed to counteract sources of cognitive bias at the level of deliberating groups. Epistemic paternalists typically focus on scaffolding the cognition of individuals, but groups can benefit no less from environmental interventions. In DPs, groupthink and the common knowledge effect are minimized by balanced briefing materials, plenary sessions with a variety of partisan experts, and the viewpoint diversity that is secured by representative selection. Trained moderators ensure that each view gets a fair hearing during group discussions, which minimizes cascades and group polarization. And because DPs are small and influential, their participants are incentivized to rationally update their beliefs, rather than retain identity-congruent convictions.

Individual training can also influence the prospects of group cognition since high functioning individuals generally make up high functioning groups. Recall Mercier and Sperber's claim that groups overwhelmingly avoid biased solutions when at least one member can effectively argue for the correct solution, and other members can be persuaded by effective arguments. There is now extensive evidence that this generally happens when groups are asked questions that admit of demonstrably correct answers, such as logical or mathematical problems (Davis 1973; Laughlin and Ellis 1986; Bonner *et al.* 2002). Given that our ability to solve such problems depends on our possessing and applying appropriate mindware, it stands to reason that groups are better off when their constituent members are trained to use a wide range of reliable mindware. Furthermore, Mercier and Sperber conjecture: "If we generalize to problems that do not have a provable solution, we should at least expect good arguments to win, even if this is not always sufficient for truth to win" (Mercier and Sperber 2011: 72). This is precisely what Tetlock found in forecasting tournaments: "Teams of ordinary forecasters beat the

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

wisdom of the crowd by about 10%. Prediction markets beat ordinary teams by about 20%. And superteams beat prediction markets by 15% to 30%" (Tetlock and Gardner 2015: 207). Superteams are made up of superforecasters who outperform the vast majority of other participants because they possess several key intellectual virtues. Furthermore, as reported above, bringing superforecasters together into groups yields a more significant improvement than grouping ordinary forecasters. This suggests that by cultivating corrective virtues in individuals, we can mitigate biased cognition in collectives. And, in turn, well structured deliberation can effectively debias the thinking of individuals. This is the sort of positive feedback loop that becomes more visible from an interactionist point of view (Skorburg 2017: 465). And once visible, these feedback loops are more likely to be productively harnessed.

Thus, what happens at the level of collectives depends on features of their constituent members, as well as the environments in which they're deliberating. More generally, because individuals, environments, and collectives are in constant interaction, the aims of virtue theoretic, paternalist, and collectivist programmes are entwined in ways that their proponents seldom acknowledge.

So far, we have examined these interactions through the lens of empirical science. But understanding the vulnerabilities of the sciences themselves to cognitive biases requires a firm grasp of the complex interactions that give rise to their results. This realization has become especially important in light of the recent discovery that an alarming number of influential studies in the life and behavioural sciences fail to replicate (Errington *et al.* 2014; Klein *et al.* 2014; OSC 2015). These findings, and their reception within the scientific community, have been widely dubbed 'the replication crisis' (or 'reproducibility crisis'). The crisis is the product not only of the apparently high rate of false positive errors in these sciences, but of the overconfidence that blinded scientists to this fact. Both of these factors are themselves partly explicable in terms of cognitive biases. In response to the replication crisis, an increasing number of practitioners are calling for a 'credibility revolution', whose many reforms defy strictly virtue theoretic, paternalist, or collectivist perspectives (Vazire 2018). A brief interactionist account of the causes of, and responses to, the replication crisis is the focus of the concluding section.

7. Interactionism and the replication crisis

My brief account of the replication crisis will recapitulate many of this paper's central themes. First, cognitive biases are very difficult to detect, identify, and mitigate, even for scientists who have been trained to do so, and even among scientists who study cognitive bias. Second, biased cognition has many sources, operating at different levels – mindware problems, cognitive miserliness, motivated reasoning, hostile environments, a lack of critical dialogue – none of which can be properly understood in isolation from the others. Finally, effectively addressing these sources of bias requires a system of coordinated interventions at the levels of individuals, situations, and collectives, i.e., inside, outside, inside-out, and outside-in strategies.

Null Hypothesis Significance Testing (NHST) is often identified as a key contributor to the replication crisis. This blend of earlier approaches to statistical reasoning is now an entrenched norm in the life and behavioural sciences (Cristea and Ioannidis 2018). According to this procedure, theoretical (alternative) hypotheses are confirmed when experimental results are inconsistent with the null hypothesis, i.e., the hypothesis that there is no relationship between the phenomena being studied. The threshold for rejecting the null hypothesis is typically $p \leq 0.05$. Of course, whether or not we should believe that the null hypothesis is false, and the alternative hypothesis is true, depends on many other factors, including the *statistical power* of the study, i.e., the probability of its

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

correctly finding an effect, if there is one. Yet, calculating the statistical power of an experiment is *not* an intrinsic component of NHST, and consequently it often goes unreported in the publication of experimental results (Gigerenzer *et al.* 2004). Worse still, the average and median statistical power of psychological studies over the last 50 years has consistently been found to be *below* 50% (Cohen 1962; Sedlmeier and Gigerenzer 1989; Szucs and Ioannidis 2017). This seems to indicate both a lack of methodological rigour on the part of authors, and an overly credulous attitude in their audiences. How can this be, when psychologists are best positioned to recognize the shortcomings of human cognition, and apparently well equipped to overcome these shortcomings by statistical reasoning? Some have argued that heuristics are to blame. Forty years before the replication crisis, Tversky and Kahneman found that professional psychologists' reliance on the representativeness heuristic results in

> the expectation that a valid hypothesis about a population will be represented by a statistically significant result in a sample with little regard for its size. As a consequence, the researchers put too much faith in the results of small samples and grossly overestimated the replicability of such results. In the actual conduct of research, this bias leads to the selection of samples of inadequate size and to overinterpretation of findings. (Tversky and Kahneman 1974: 1126)

Gigerenzer, on the other hand, blames recent trends in statistical training. He argues that scientists are trained to treat NHST as a mechanical, unreflective decision procedure for the evaluation of all experimental results. The resultant *ritualistic* use of NHST has given rise to delusions that blind scientists to its limitations: scientists confuse *p*-values with the probability that a result will replicate (replication delusion);[8] they believe that statistically significant results constitute absolutely definitive evidence (illusion of certainty); they confuse *p*-values for the probability that the null hypothesis is true (inverse probability error); and, as mentioned, they pay little attention to statistical power (Gigerenzer 2018). In any case, whether one points the finger at the representativeness heuristic or an uncritical overreliance on NHST (or both), it's clear that bad mindware – *reliabilist vices* – are partly responsible for both the high rate of false positive errors and the overconfidence that gave rise to the replication crisis.

NHST has also been abused by scientists who use questionable research practices to enhance their experimental findings. These practices include: *p*-hacking; cherry-picking; and hypothesizing after the results are known (HARKing). *P*-hacking is an attempt to manipulate experimental data to get statistically significant results.[9] Scientists do this by rounding-down *p*-values, collecting supplementary data, and excluding data points *post hoc*. Scientists cherry-pick when they selectively exclude non-significant outcomes, such as dependent measures, from their reporting of experimental results. And HARKing includes presenting unexpected results as having been predicted by initial hypotheses, thus passing off exploratory research as confirmatory research. Large surveys indicate that a *majority* of scientists in the life and behavioural sciences have engaged in *at least* one of these practices over their careers, which has undoubtedly contributed to the false positive and overconfidence problems in those disciplines (John *et al.* 2012; Fraser *et al.* 2018). This trend supports Kahan's claim that inferentially and reflectively sophisticated agents are best able to use their cognitive resources to satisfy

[8]For example, if a *p*-value of 0.03 is obtained, some scientists think that the probability of replicating the experimental result is 0.97 (Rosenthal 1993).

[9]*P*-hacking, and the other questionable research practices, are distinguished from scientific fraud insofar as they are not conscious attempts to deceive a study's audience.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

their non-epistemic motivations, which often leads them to more biased results. The questionable research practices that scientists engage in when doing so seem best classified as *responsibilist vices*.

The replication crisis cannot be laid solely at the feet of practicing scientists, however. Its causes extend into the scientific *establishment* as well, i.e., the set of institutional, social, and informational *environments* in which scientific inquiry is conducted, reported, and evaluated. Granting agencies and academic journals express, implicitly and explicitly, a strong preference to fund and publish positive, original findings (Schmidt 2009; Fanelli 2012). This longstanding publication bias against inconclusive and replication studies has created a file-drawer problem: the scientific literature likely does not accurately reflect the state of evidence for its positive claims because negative evidence does not get publicized (Sterling 1959). Consequently, meta-analyses often cannot be trusted to fulfil their epistemic purpose of increasing the power and/or precision of scientific findings (Kicinski 2013). Furthermore, without open channels of *critical* dialogue, the sciences cannot effectively harness the bias mitigating forces of dialectical argumentation.

Any account of the biases responsible for the replication crisis that focuses on these personal, situational, and collectivist factors in isolation from one another is bound to be incomplete and misleading, for they are just as much the result of *interactions* between these factors as they are to their combination. For example, overreliance on NHST has played a role in generating publication bias: the threshold for statistical significance gives editors a target that too often informs their decisions about what to publish (Meehl 1967, 1978). In turn, publication bias, together with the publish-or-perish climate of academia, motivates questionable research practices: in a highly competitive environment, where status is largely determined by the number and impact of one's publications, practitioners are highly motivated to present their findings as both statistically and theoretically significant (Fanelli 2010). And research practices that focus overwhelmingly on achieving statistical significance reinforce the ritualistic use of NHST (Gigerenzer 2018). Thus, effective responses to the replication crisis must be multi-faceted and well integrated, in a way that virtue theoretic, paternalist, and collectivist responses tend not to be: they must address epistemic shortcomings at each level of analysis, and do so in an outward-looking way that increases our chances of making improvements at other levels. What's required are inside, outside, inside-out, and outside-in strategies that promote epistemic synergy. Fortunately, there are many such strategies on offer.

Inside strategies generally have the goal of replacing contaminated mindware – NHST, as it is currently practiced – with more reliable alternatives.[10] To this end, some have argued that scientists should adopt a stricter standard for rejecting the null hypothesis by decreasing the threshold for statistical significance (Benjamin *et al.* 2018). Others have argued that the threshold should be variable, context-sensitive, and subject to argumentation (Lakens *et al.* 2018). On the other hand, Gigerenzer insists that any strategy that focuses on the threshold of statistical significance does nothing to address the ritualistic use of NHST, with its attendant delusions. Instead, he claims that statistical training should impart a multiplicity of methods – what he calls 'the statistical toolbox' – as well as the good judgement to apply them when appropriate:

[10]Questionable research practices, on the other hand, are often addressed by outside strategies. This, it seems, is an efficient division of cognitive labour given the difficulties inherent in cultivating the corrective responsibilist virtues through education and training. A more elaborate argument for this policy can be found in Bland (Forthcoming).

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

> Psychology departments need to begin teaching the statistical toolbox. This toolbox includes techniques to visualize the descriptive statistics of the sample, Tukey's exploratory data analysis, meta-analysis, estimation, Fisher's null-hypothesis testing (which is not the same as the null ritual), Neyman-Pearson decision theory, and Bayesian inference. Most important, the toolbox approach requires using informed judgment to select the appropriate tool for a given problem. (Gigerenzer 2018: 213)

I do not wish to take a stand on these controversial matters. Rather, I wish only to frame these strategies as having the common aim of developing *reliabilist virtues* that are meant to counteract some of the biases at the heart of the replication crisis.

Many scientific journals have adopted strategies of addressing some of the external factors responsible for the replication crisis. Among them is the policy, adopted by more than 200 journals to date, of basing publication decisions on blind peer reviews of registered reports (Center for Open Science 2020*b*). Registered reports include introduction, methods, and planned analysis sections in anticipation of an experiment that the author (s) has yet to carry out. When the reports are deemed acceptable by reviewers, the journal pledges to publish the results of the experiment, regardless of how they turn out (as long as the experiment is carried out as planned, and within a specified timeframe). This policy has the dual benefit of curbing publication bias, and decreasing the motivation for scientists to engage in questionable research practices. As such, it is both a paternalistic, outside strategy of improving the informational environment within scientific communities, and an outside-in strategy of fostering responsibilist virtues (and/or attenuating responsibilist vices). Many journals are also taking measures to encourage the practice of open science, which is "the process of making the content and process of producing evidence and claims transparent and accessible to others" (Munafò *et al.* 2017: 5). One such measure is the adoption of Transparency and Openness Promotion (TOP) Guidelines for the reporting of scientific research. The TOP Guidelines encourage authors to disclose not only the results of their research, but the analytical methods (code), research design, and data that led them to their conclusions (Nosek *et al.* 2015). In doing so, they give scientists an opportunity to more thoroughly evaluate one another's research, and create a more fertile environment for critical dialogue within the scientific community. More than 1000 journals have implemented this outside strategy of harnessing the bias-mitigating forces of collective deliberation (Center for Open Science 2020*a*).

Finally, there are steps that scientists themselves have taken to improve the informational environments in which their inquiries take place. One example is the practice of pre-registration, which consists in making a public prediction about an experiment's results *before* the data are collected, viewed, or analyzed. This self-imposed constraint mitigates hindsight bias, helps to distinguish confirmatory research from exploratory research (Nosek *et al.* 2018), and mitigates HARKing (Kerr 1998). It is an inside-out strategy of minimizing the reliabilist and responsibilist vices that give rise to biased cognition.

At present it is difficult to say which of these strategies, or others, will be most successful in ameliorating the biases responsible for the replication crisis; they must be tested diachronically. Furthermore, they must be evaluated *holistically*, since each strategy's efficacy depends on which other strategies are adopted, and how they're implemented. For example, the critical dialogue fostered by norms of open science would plausibly be improved by the widespread use of reliable methods of statistical inference. And more scientists may be encouraged to pre-register their hypotheses if a sufficiently large number of journals use registered reports to make their editorial decisions.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

Without appreciating these interdependencies, we stand little chance of coordinating strategies in ways that most productively curb biased cognition in the life and behavioural sciences. The first step in doing so is to adopt an interactionist view, rather than an isolationist one along virtue theoretic, paternalist, or collectivist lines.[11]

References

Ahlstrom-Vij K. (2013). *Epistemic Paternalism: A Defense*. New York, NY: Palgrave Macmillan.

Alfano M. (2016). 'Friendship and the Structure of Trust.' In A. Masala and J. Webber (eds), *From Personality to Virtue: Essays in the Psychology and Ethics of Character*, pp. 186–206. Oxford: Oxford University Press.

Armor D.A. and Taylor S.E. (2002). 'When Predictions Fail: The Dilemma of Unrealistic Optimism.' In T. Gilovich, D. Griffin and D. Kahneman (eds), *Heuristics and Biases: The Psychology of Intuitive Judgment*, pp. 334–47. Cambridge: Cambridge University Press.

Augustinova M. (2008). 'Falsification Cueing in Collective Reasoning: Example of the Wason Selection Task.' *European Journal of Social Psychology* **38**(5), 770–85.

Baron R.S., Hoppe S.I., Kao C.F., Brunsman B., Linneweh B. and Rogers, D. (1996). 'Social Corroboration and Opinion Extremity.' *Journal of Experimental Social Psychology* **32**, 537–60.

Benjamin D.J., Berger J.O., Johannesson M. *et al.* (2018). 'Redefine Statistical Significance.' *Nature Human Behaviour* **2**(1), 6–10.

Bland S. (Forthcoming). 'Interactionism, Debiasing, and the Division of Epistemic Labour.' Forthcoming in M. Alfano, C. Klein and J. De Ridder (eds), *Social Virtue Epistemology*. London: Routledge.

Bonner S.E., Baumann M.R. and Dalal R.S. (2002). 'The Effects of Member Expertise on Group Decision Making and Performance.' *Organizational Behavior and Human Decision Processes* **88**, 719–36.

Cassam Q. (2015). 'Stealthy Vices.' *Social Epistemology Review and Reply Collective* **4**, 19–25.

Casscells W., Schoenberger A. and Grayboys T. (1978). 'Interpretation by Physicians of Clinical Laboratory Results.' *New England Journal of Medicine* **299**, 999–1001.

Center for Open Science (2020a). 'TOP Guidelines.' https://cos.io/top/.

Center for Open Science (2020b). 'Registered Reports.' https://cos.io/rr/.

Cohen J. (1962). 'The Statistical Power of Abnormal-Social Psychological Research: A Review.' *Journal of Abnormal and Social Psychology* **65**(3), 145–53.

Cristea I.A. and Ioannidis J.P.A. (2018). 'P-values in Display Items are Ubiquitous and Almost Invariably Significant: A Survey of Top Science Journals.' *PLoS ONE* **13**(5), e0197440. doi:10.1371/journal.pone.0197440.

Davis J.H. (1973). 'Group Decisions and Social Interactions: A Theory of Social Decision Schemes.' *Psychological Review* **80**(2), 97–125.

Dunning D., Meyerowitz J.A. and Holzberg A.D. (2002). 'Ambiguity and Self-evaluation: The Role of Idiosyncratic Trait Definition in Self-serving Assessments of Ability.' In T. Gilovich, D. Griffin and D. Kahneman (eds), *Heuristics and Biases: The Psychology of Intuitive Judgment*, pp. 324–33. Cambridge: Cambridge University Press.

Ehrlinger J., Gilovich T. and Ross L. (2005). 'Peering into the Bias Blind Spot: Peoples' Assessments of Bias in Themselves and Others.' *Personality and Social Psychology Bulletin* **31**(5), 680–92.

Errington T.M., Iorns E., Gunn W.T., Fraser E., Lomax J. and Nosek B.A. (2014). 'An Open Investigation of the Reproducibility of Cancer Biology Research.' *ELife* **3**(December), e043333. doi: 10.7554/eLife.04333.

Evans J.St.B.T. (2011). 'Reason is for Thinking, Not Just for Arguing.' *Behavioral and Brain Sciences* **34**(2), 77–9.

Fanelli D. (2010). 'Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data.' *PLoS ONE* **5**(4). doi: 10.1371/journal.pone.0010271.

Fanelli D. (2012). 'Negative Results are Disappearing from Most Disciplines and Countries.' *Scientometrics* **90**(3), 891–904.

Fishkin J.S. (2009). *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford: Oxford University Press.

[11]I am grateful to Mark Alfano, Amy Bell, Jon Marsh, and an anonymous referee for very helpful comments on earlier drafts of this paper.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

Fong G.T., Krantz D.H. and Nisbett R. (1986). 'The Effects of Statistical Training on Thinking about Everyday Problems.' *Cognitive Psychology* **18**, 253–92.

Frankish K. (2011). 'Reasoning, Argumentation, and Cognition.' *Behavioral and Brain Sciences* **34**(2), 79–80.

Frantz C. (2006). 'I AM Being Fair: The Bias Blind Spot as a Stumbling Block to Seeing Both Sides.' *Basic and Applied Social Psychology* **28**(2), 157–67.

Fraser H., Parker T., Nakagawa S., Barnett A. and Fidler F. (2018). 'Questionable Research Practices in Ecology and Evolution.' *PLoS ONE*, **13**(7). doi: 10.1371/journal.pone.0200303.

Gigerenzer G. (1991). 'How to Make Cognitive Illusions Disappear.' *European Review of Social Psychology* **2**, 83–115.

Gigerenzer G. (1996). 'The Psychology of Good Judgment: Frequency Formats and Simple Algorithms.' *Medical Decision Making* **16**, 273–80.

Gigerenzer G. (2018). 'Statistical Rituals: The Replication Delusion and How We Got There.' *Advances in Methods and Practices in Psychological Science* **1**(2), 198–218.

Gigerenzer G. and Hoffrage U. (1995). 'How to Improve Bayesian Reasoning Without Instruction: Frequency Formats.' *Psychological Review* **102**, 684–704.

Gigerenzer G., Todd P.M. and the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. New York, NY: Oxford University Press.

Gigerenzer G., Kraus S. and Vitouch O. (2004). 'The Null Ritual: What You Always Wanted to Know about Null Hypothesis Testing but Were Afraid to Ask.' In D. Kaplan (ed.), *Handbook on Quantitative Methods in the Social Sciences*, pp. 391–408. Thousand Oaks, CA: Sage.

Gigone D. and Hastie R. (1993). 'The Common Knowledge Effect: Information Sharing and Group Judgements.' *Journal of Personality and Social Psychology* **56**, 959–74.

Habermas J. (1990). *Moral Consciousness and Communicative Action*. Cambridge, MA: MIT Press.

Hertwig R. and Gigerenzer G. (1999). 'The "Conjunction Fallacy" Revisited: How Intelligent Inferences Look Like Reasoning Error.' *Journal of Behavioral Decision Making* **12**, 275–305.

John L.K., Loewenstein G. and Prelec D. (2012). 'Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling.' *Psychological Science* **23**(5), 524–32.

Kahan D.M. (2013). 'Ideaology, Motivated Reasoning, and Cognitive Reflection.' *Judgment and Decision Making* **8**, 407–24.

Kahan D.M., Peters E., Dawson E.C. and Slovic P. (2017). 'Motivated Numeracy and Enlightened Self-government.' *Behavioural Public Policy* **1**(1), 54–86.

Kahneman D. (2011). *Thinking Fast and Slow*. London: Penguin Books.

Kenyon T. and Beaulac G. (2014). 'Critical Thinking Education and Debiasing.' *Informal Logic* **34**(4), 341–63.

Kerr N.L. (1998). 'HARKing: Hypothesizing After the Results are Known.' *Personality and Social Psychology Review* **2**(3), 196–217.

Kicinski M. (2013). 'Publication Bias in Recent Meta-analyses.' *PLoS ONE* **8**(11). doi: 10.1371/journal.pone.0081823.

Kihlstrom J. (2013). 'The Person-situation Interaction.' In D. Carlston (ed.), *The Oxford Handbook of Social Cognition*, pp. 786–806. Oxford: Oxford University Press.

Klein R.A., Ratliff K.A., Vianello M. *et al.* (2014). 'Investigating Variation in Replicability: A 'Many Labs' Replication Project.' *Social Psychology* **45**(3), 142–52.

Kunda Z. (1990). 'The Case for Motivated Reasoning.' *Psychological Bulletin* **108**(3), 480–98.

Lakens D., Adolfi F.G., Albers C.J. *et al.* (2018). 'Justify your Alpha.' *Nature Human Behaviour* **2**(3), 168–71.

Laughlin P.R. and Ellis A.L. (1986). 'Demonstrability and Social Combination Processes on Mathematical Intellective Tasks.' *Journal of Experimental Social Psychology* **22**(3), 177–89.

Lehman D.R., Lempert R.O. and Nisbett R.E. (1988). 'The Effects of Graduate Training on Reasoning: Formal Discipline and Thinking about Everyday-life Events.' *American Psychologist* **43**(6), 431–42.

Lehman D.R. and Nisbett R.E. (1990). 'A Longitudinal Study of the Effect of Undergraduate Training on Reasoning.' *Developmental Psychology* **26**(6), 952–60.

Longino H. (1990). *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Lorenz J., Rauhut H. and Schweitzer F. (2011). 'How Social Influences can Undermine the Wisdom of Crowd Effect.' *Proceedings of the National Academy of Sciences USA* **108**, 9020–5.

Meehl P.E. (1967). 'Theory-testing in Psychology and Physics: A Methodological Paradox.' *Philosophy of Science* **34**(2), 103–15.

Meehl P.E. (1978). 'Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology.' *Journal of Consulting and Clinical Psychology* **46**(4), 806–34.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

Mercier H. and Sperber D. (2011). 'Why do Humans Reason? Arguments for an Argumentative Theory.' *Behavioral and Brain Sciences* **34**(2), 57–74.

Mercier H. and Sperber D. (2017). *The Enigma of Reason*. Cambridge, MA: Harvard University Press.

Mill J.S. (2003 [1859]). *On Liberty*. New Haven, CT: Yale University Press.

Moshman D. and Geil M. (1998). 'Collaborative Reasoning: Evidence for Collective Rationality.' *Thinking and Reasoning* **4**(3), 231–48.

Munafò M.R., Nosek B.A., Bishop D.V.M., Button K.S., Chambers C.D., Percie du Sert N., Simonsohn U., Wagenmakers E.J., Ware J.J. and Ioannidis J.P.A. (2017). 'A Manifesto for Reproducible Science.' *Nature Human Behaviour* **1**(1), 0021. doi: 10.1038/s41562-016-0021.

Nisbett R.E. and Wilson T.D. (1977). 'Telling More than We can Know: Verbal Reports on Mental Processes.' *Psychological Review* **84**(3), 231–59.

Nosek B.A., Alter G., Bank G.C. *et al.* (2015). 'Promoting an Open Research Culture.' *Science* **348**, 1422–5.

Nosek B.A., Ebersole C.R., DeHaven A.C. and Mellor D.T. (2018). 'The Preregistration Revolution.' *Proceedings of the National Academy of Sciences USA* **115**(11), 2600–6.

Open Science Collaboration (2015). 'Estimating the Reproducibility of Psychological Science.' *Science* **349** (6251), 943–51.

Popper K. (2002 [1996]). *The Open Society and Its Enemies*. Fifth Edition. New York, NY: Routledge.

Pronin E. (2007). 'Perception and Misperception of Bias in Human Judgment.' *Trends in Cognitive Sciences* **11**(1), 37–43.

Pronin E., Lin D. and Ross L. (2002). 'The Bias Blind Spot: Perceptions of Bias is Self Versus Others.' *Personality and Social Psychology Bulletin* **28**, 369–81.

Pronin, E., Gilovich, T. and Ross, L. (2004). Objectivity in the Eye of the Beholder: Divergent Perceptions of Bias in Self Versus Others. *Psychological Review* **111**, 781–99.

Ritchart R. and Perkins D.N. (2005). 'Learning to Think: The Challenges of Teaching Thinking.' In K. Holyoak and R. Morrison (eds), *The Cambridge Handbook of Thinking and Reasoning*, pp. 775–802. Cambridge: Cambridge University Press.

Roberts, R.C. and West, R. (2015). 'Natural Epistemic Defects and Corrective Virtues.' *Synthese* **192**, 2557–76.

Rosenthal R. (1993). 'Cumulating Evidence.' In G. Keren and C. Lewis (eds), *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues*, pp. 519–59. Hillsdale, NJ: Lawrence Erlbaum.

Samuelson P.L. and Church I.M. (2015). 'When Cognition Turns Vicious: Heuristics and Biases in Light of Virtue Epistemology.' *Philosophical Psychology* **28**(8), 1095–113.

Sanna L., Stocker S. and Schwarz N. (2002). 'When Debiasing Backfires: Accessible Content and Accessibility Experiences in Debiasing Hindsight.' *Journal of Experimental Psychology: Learning, Memory, and Cognition* **28**(3), 497–502.

Schmidt S. (2009). 'Shall We Really do it Again? The Powerful Concept of Replication is Neglected in the Social Sciences.' *Review of General Psychology* **13**(2), 90–100.

Sedikides C., Horton R.S. and Gregg A.P. (2007). 'The Why's The Limit: Curtailing Self-enhancement with Explanatory Introspection.' *Journal of Personality* **75**, 783–824.

Sedlmeier P. and Gigerenzer G. (1989). 'Do Studies of Statistical Power Have an Effect on the Power of Studies?' *Psychological Bulletin* **105**, 309–16.

Skorburg J.A. (2017). 'Lessons and New Directions for Extended Cognition from Social and Personality Psychology.' *Philosophical Psychology* **30**(4), 458–80.

Skorburg J.A. and Alfano M. (2019). 'Psychological Science and Virtue Epistemology: Intelligence as an Interactionist Virtue.' In H. Battaly (ed.), *The Routledge Handbook of Virtue Epistemology*, pp. 433–45. New York, NY: Routledge.

Stanovich K.E. (2011). *Rationality and the Reflective Mind*. Oxford: Oxford University Press.

Sterling T.D. (1959). 'Publication Decisions and their Possible Effects on Inferences Drawn from Tests of Significance – or Vice Versa.' *Journal of the American Statistical Association* **54**(285), 30–4.

Sullivan E., Sondag M., Rutter I., Meulemans W., Cunningham S., Speckmann B. and Alfano M. (Forthcoming). 'Can Real Social Epistemic Networks Deliver the Wisdom of Crowds?' In T. Lambrozo, J. Knobe and S. Nichols (eds), *Oxford Studies in Experimental Philosophy*. Oxford: Oxford University Press.

Sunstein C. (2003). 'The Law of Group Polarization.' In J.S. Fishkin and P. Laslett (eds), *Debating Deliberative Democracy*, pp. 80–101. Oxford: Blackwell.

Sunstein C. (2009). *Going to Extremes: How Like Minds Unite and Divide*. Oxford: Oxford University Press.

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press

Szucs D. and Ionnidis J.P.A. (2017). 'Empirical Assessment of Published Effect Sizes and Power in the Recent Cognitive Neuroscience and Psychological Literature.' *PLOS Biology* **15**(3). doi: 10.1371/journalpbio.2000797.

Tetlock P. (2005). *Expert Political Judgment*. Princeton, NJ: Princeton University Press.

Tetlock P. and Gardner D. (2015). *Superforecasting: The Art and Science of Prediction*. Toronto: Signal.

Trout J.D. (2005). 'Paternalism and Cognitive Bias.' *Law and Philosophy* **24**, 393–434.

Tversky A. and Kahneman D. (1974). 'Judgement under Uncertainty: Heuristics and Biases.' *Science* **185** (4157), 1124–31.

Vazire S. (2018). 'Implications of the Credibility Revolution for Productivity, Creativity, and Progress.' *Perspectives on Psychological Science* **13**(4), 411–17.

Willingham D.T. (2008). 'Critical Thinking: Why is it so Hard to Teach?' *Arts Education Policy Review* **109** (4), 21–32.

Wilson T.D. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.

Wilson T.D., Centerbar D.B. and Brekke N. (2002). 'Mental Contamination and the Debiasing Problem.' In T. Gilovich, D. Griffin and D. Kahneman (eds), *Heuristics and Biases: The Psychology of Intuitive Judgment*, pp. 185–200. Cambridge: Cambridge University Press.

Steven Bland is an Associate Professor of Philosophy at Huron University College. His work focuses on the intersection of epistemology and philosophy of science, and has been published in *Synthese*, *Episteme*, *Social Epistemology*, and *Studies in History and Philosophy of Science*. His book, *Epistemic Relativism and Scepticism: Unwinding the Braid*, was published by Palgrave in 2018.

Cite this article: Bland S (2022). An Interactionist Approach to Cognitive Debiasing. *Episteme* **19**, 66–88. https://doi.org/10.1017/epi.2020.9

https://doi.org/10.1017/epi.2020.9 Published online by Cambridge University Press