Brit. J. Phil. Sci. **0** (2020), 1–24

Illegitimate Values, Confirmation Bias, and Mandevillian Cognition in Science

Uwe Peters

ABSTRACT

The proposal that values in science are illegitimate and that they should be counteracted whenever they direct inquiry to the confirmation of predetermined conclusions is not uncommon in the philosophy of science. Drawing on recent research from cognitive science on human reasoning and confirmation bias, I argue that this view should be rejected. Values that drive inquiry to the confirmation of predetermined conclusions can contribute to the reliability of scientific inquiry at the group level, even when they negatively affect an individual's cognition. This casts doubt on the proposal that such values should always be illegitimate in science. It also suggests that this proposal assumes a narrow, individualistic account of science that threatens to undermine the project of ensuring reliable belief-formation in science.

doi:10.1093/bjps/axy079

1 Introduction

Science involves different kinds of values. Cognitive and non-cognitive values are often distinguished (Longino [1996]; Douglas [2013]).[1] Cognitive values include truth, empirical adequacy, consistency, simplicity, fruitfulness, and explanatory power. They are taken to be legitimate in and constitutive of science (Lacey [1997]). I shall here set them aside.

I want to focus on non-cognitive values. Non-cognitive values are, for example, moral, prudential, political, and aesthetic values. It is now widely accepted that they too may play legitimate roles in science. They are taken to be acceptable, for instance, as reasons to investigate particular scientific problems and endorse certain conceptualizations (Alexandrova [2018]), as ethical constraints on scientific studies and research protocols (Elliot [2017]), as arbiters between underdetermined theories (Longino [2002]), or as determinants of standards of confirmation (Douglas [2009]).

They might, however, also pose problems in the sciences. As Anderson ([2004], p. 2) notes:

> Yet surely some uses of values [in science] to select background assumptions are illegitimate. Feminists object to the deployment of sexist values to select background assumptions that insulate the theoretical underpinnings of patriarchy from refutation. Critics of feminist science similarly worry that feminists will use their values in ways that insulate feminist theories from refutation. We need criteria to distinguish legitimate from illegitimate ways of deploying values in science.

Many philosophers have written on the question of how we should distinguish legitimate from illegitimate uses of values in science (Anderson [2004]; Douglas [2009]; Hicks [2014]; Intemann [2015]; Elliot [2017]). In this article I want to assess the tenability of one common criterion used to draw the distinction, namely, the view that values are illegitimate in science and their influence should be counteracted when they drive inquiry to the confirmation of favoured, predetermined conclusions. I shall refer to values that have this functional profile as ‘confirmatory values’, and I shall call the view at issue the ‘confirmatory value’ (CV) view.

The CV view is widely accepted in the philosophical literature on values in science (Anderson [2004]; Brown [2013]; Douglas [2016]; De Melo-Martin and Intemann [2016]; Elliott [2017]), and it is *prima facie* highly plausible. For it seems clear that, in contrast to scientists impartially assessing evidence for and against all claims, when values impel scientists to corroborate already endorsed claims, this one-sided information processing threatens the reliability of belief-formation in science.

[1] This is not to say that philosophers working on values in science generally endorse this distinction; some are critical of it (Rooney [1992]; Longino [1996]).

But are confirmatory values always epistemically problematic in science, and is the CV view in its generality tenable? The question is important, because our answer to it is directly relevant to how science should be done (namely, with or without confirmatory values).

I shall argue against the CV view. I will do so by discussing recent cognitive scientific research on human reasoning and confirmation bias. This research indicates that even though confirmation bias is epistemically detrimental for individual reasoners, it can be epistemically beneficial for groups of them (Mercier and Sperber [2011], [2017]).

Epistemically imperfect mental processes or states that have such group-level benefits have been called 'Mandevillian' cognitions (Morton [2014]; Smart [2018]), after Bernard Mandeville ([1705]), who was the first to propose that an individual's private cognitive and moral shortcomings may promote public goods.[2] The implications of Mandevillian cognition in general, and confirmation bias in particular, have so far not been explored in the context of the debate about how to distinguish legitimate from illegitimate values in science.

This is unfortunate because, as I shall argue, Mandevillian aspects of confirmation bias in scientific inquiry suggest that confirmatory values too can be epistemically beneficial, contributing to the reliability of science at the group level, even though they negatively affect an individual's cognition. This casts doubt on the proposal that such values should always be illegitimate in science. Moreover, it suggests that advocates of the CV view assume a narrow, individualistic account of scientific inquiry that threatens to undermine their own project of ensuring reliable belief-formation in science.

In Sections 2 and 3, I provide textual evidence of the CV view in the debate on values in science, specifiy the version of the view that I will focus on, and outline my argumentative strategy to assess it. In Section 4, I introduce research on human reasoning and the Mandevillian character of confirmation bias in science. In Section 5, I use that research to argue against the CV view, qualify that argument, and rebut an objection to it. Section 6 summarizes and concludes the discussion.

2 Advocates of the Confirmatory Value View

Given the *prima facie* plausibility of the proposal that values (or value judgements[3]) are illegitimate in science when they direct inquiry to pre-existing

[2] In his fable *The Grumbling Hive*, Mandeville ([1705]) wrote (*inter alia*): 'every part [of the hive] was full of vice, yet the whole mass a paradise'.

[3] Values are not value judgements, but the difference does not matter here and the two can be treated interchangeably.

conclusions, it is not surprising that many philosophers of science endorse the CV view. For instance, Anderson ([2004], p. 11) holds:

> We need to ensure that value judgments do not operate to drive inquiry to a predetermined conclusion. This is our fundamental criterion for distinguishing legitimate from illegitimate uses of values in science.

We need to make sure, Anderson continues, that the 'evaluative presuppositions brought to inquiry do not determine the answer to the evaluative question in advance, but leave this open to determination by the evidence'. 'If a hypothesis is to be tested, the research design must leave open a fair possibility that evidence will disconfirm it' rather than direct scientists towards its confirmation (Anderson [2004], p. 19). These comments suggest that Anderson endorses the CV view.

Douglas ([2016], p. 618) seems to subscribe to it too, writing:

> Most problematically, values in a direct role during evidential assessment would be equivalent to allowing wishful thinking into the heart of science. If values could play a direct role in the assessment of evidence, a preference for a particular outcome could act as a reason for that outcome or for the rejection of a disliked outcome.

And this, Douglas ([2016], p. 618) holds, is 'unacceptable'.

Similarly, she maintains that while values might play a legitimate role in the early phases of science, for instance, in the selection of research topics and methodologies:

> One cannot use values to direct the selection of a problem and a formulation of a methodology that in combination predetermines (or substantially restricts) the outcome of a study. Such an approach undermines the core value of science—to produce reliable knowledge—which requires the possibility that the evidence produced could come out against one's favoured theory. (Douglas [2009], p. 100)

When values play a direct role in evidential assessment or in the choice of a methodology (that corroborates a favoured view), values are illegitimate for Douglas because they incline scientists to accept (or reject) a particular conclusion on the basis of a preference for (or aversion against) it, rather than on the basis of the evidence alone. Via their involvement in the assessment of evidence or in the choice of methodology, values may skew inquiry and direct it to pre-existing, preference-based outcomes. Thus Douglas ([2009], [2016]) too endorses the CV view.

Other philosophers follow suit. For instance, Brown ([2013], p. 835) writes that the 'main concern' about values in science is that 'value judgments might "drive inquiry to a predetermined conclusion"', leading 'inquirers [to] rig the game in favour of their preferred values'. The 'key to the problem' posed by values in science, Brown ([2013], p. 838) adds, is to ensure that we do 'not predetermine the conclusion of inquiry, that we leave ourselves open to surprise'. Elliot ([2017], p. 13) agrees, writing that 'values [are] unacceptable [in science when they lead to practices such as] ignoring evidence that conflicts with one's preferred conclusions [and] using "rigged" methods that generate predetermined outcomes'.

Even philosophers who hold that objectivity is not a property of an individual but of a group, and who maintain that individuals' preferences and values can be epistemically beneficial for the group as a whole (for example, in sustaining intellectual diversity) still tend to wish to control the influence of preferences and values in science in ways that suggest an endorsement of the CV view. For instance, Longino ([1990], [2002]) argues that objectivity is not to be found in individual scientists since their cognition is limited and affected by subjective idiosyncrasies. Rather, objectivity results from social interactions involving an extensive and comprehensive mix of different subjective preferences and values that cancel each other out in a process of social criticism (Longino [1990], p. 73).

Crucially, on Longino's view, for social criticism to be able to 'limit' the 'intrusion [of] subjective preferences' in science, individual scientists must not be driven to the confirmation of favoured, predetermined conclusions but need to 'take up', and be responsive to, critical social feedback, leaving their conclusions open to it (Longino [1990], p. 78, [2002], p. 130). That is, Longino too views subjective preferences, which include confirmatory values, as epistemically detrimental to science and calls for them to be kept in check by each scientist's adherence to this 'uptake' condition.

It is fair to say, then, that many if not most philosophers in the debate on values in science accept the CV view (for further examples, see Haack [2003]; De Melo-Martin and Intemann [2016]). There are, however, different versions of the latter. It will be useful to consider some of them before specifying which version is relevant here.

3 Versions of the Confirmatory Value View

The CV view can take different forms for at least three reasons. First, confirmatory values might direct the inquiry of an individual, a group, or both to predetermined conclusions. Relatedly, due to social interaction effects, these values might negatively affect the outcome of an individual's cognition without negatively affecting the outcome of the group's cognition, or *vice versa*. Depending on how we specify the effect of confirmatory values, we arrive at different versions of the CV view.

Second, the influence of confirmatory values on cognition comes in degrees (Wilholt [2009]). For example, they might lead an individual, a group, or both to (i) intentionally manipulate methods of collecting and assessing data so that the findings support their favoured, pre-existing conclusions.[4] Or they might lead to the unintentional adoption of methods for collecting and assessing data that are (ii) significantly skewed towards confirming such conclusions, (iii) somewhat skewed towards them, or (iv) only slightly skewed towards them. Again, depending on how we construe the influence of confirmatory values on cognition, different versions of the CV view result.

Finally, the CV view might be interpreted to apply to all cases in which confirmatory values affect cognition in science. Or it might be taken to hold only for some cases.

I have no objection to the proposal that values that drive group inquiries to predetermined conclusions are epistemically detrimental and should be illegitimate in science. I shall also not object to the view that sometimes, perhaps frequently, values affecting an individual's and/or group's inquiry in the ways described in (i)–(iv) are epistemically problematic and should be illegitimate.

The version of the CV view that is the target here is different and more general. It says that whenever an individual or group processes information unfairly as a result of values directing their inquiry to a predetermined conclusion, then these values are illegitimate in science and should be counteracted because they threaten to undermine the 'core value of science [the production of] reliable knowledge' (Douglas [2009], p. 100).

The passages cited in Section 2 suggest that, for instance, Anderson ([2004]), Douglas ([2009], [2016]), Brown ([2013]), Elliot ([2017]), and Longino ([1990], [2002]) endorse this general and, at first glance, highly plausible version of the CV view. That is not to say that they have explicitly argued for it. Rather, their comments on illegitimate values are in line with an acceptance of this view, and they have so far not attended to the distinctions just drawn, nor clarified that they endorse only a more restricted variant of it.

In what follows, I shall take this general version of the CV view to be the sole referent of the term 'CV view'. The project here is to investigate whether this view is tenable. Do values, when they drive inquiry to predetermined conclusions, always undermine the reliability of belief-formation?

The answer is not obvious. In some cases, confirmatory values might incline subjects to confirm predetermined conclusions that are in fact true. It is not clear that in such cases, these values are epistemically detrimental. After all, they incline subjects toward supporting correct claims and lead them more swiftly to the truth than a more critical mindset would, because they dispose subjects to ignore contradictory considerations. To settle whether values that

[4] I use the term 'data' broadly to refer to empirical evidence, theoretical considerations, and arguments.

direct inquiry to predetermined conclusions are always epistemically pernicious, and so illegitimate in science, thus requires further argument.

As noted, many philosophers seem to assume that these values are indeed always problematic. I shall argue that this assumption is mistaken, even if we set aside instances in which confirmatory values happen to move scientists toward truths. I want to make the point by examining the CV view in light of research on Mandevillian cognition.

4 Mandevillian Cognition and Why It Matters in Science

In everyday and scientific reasoning, we are sometimes affected by less-than-admirable epistemic states such as nosiness, obsessiveness, denial, partisanship, and various sorts of cognitive and social biases (Kahneman [2011]; Morton [2014]; Peters [2016], [forthcoming]). While it is well known that our individual judgement and decision-making is often sub-optimal as a result, some social epistemologists have explored the possibility that cognitive factors that are epistemically problematic at the individual level of information processing may be conducive to epistemic success at the group level (Kitcher [1990]; Solomon [1992]; Rowbottom [2011]).

For instance, Morton ([2014]) argues that while nosiness, obsessiveness, and denial tend to be epistemically problematic in individuals, they can have desirable epistemic effects in groups. Morton ([2014], p. 163) calls this a 'Mandevillian' effect, as he sees the idea already nascent in (Mandeville [1705]). Developing Morton's line of thought further, Smart ([2018]) offers an interesting overview of a range of cognitive phenomena that he conceptualizes as instances of 'Mandevillian intelligence'.

So far, the implications of this epistemological research on Mandevillian cognition for the normative theorizing in the philosophy of science on values in general, and the CV view in particular, have not been investigated. I want to change this. I shall do so by drawing on cognitive scientific research on a psychological phenomenon that corresponds to the functional profile of confirmatory values, namely, 'confirmation bias' (Nickerson [1998]; or 'myside bias', Stanovich *et al.* [2013]; Mercier and Sperber [2017]). Confirmation bias is typically taken to be the tendency to search for information that supports one's own pre-existing views and to ignore or distort evidence or arguments that contradict them (Myers and De Wall [2015], p. 357; Nickerson [1998]).

Confirmation bias and confirmatory values aren't the same. For instance, for some scientists, social justice and equality are political values that might also be confirmatory values. They are when they underlie a scientist's judgement and decision-making in the way outlined above. In contrast, confirmation bias is not itself a value, but rather a cognitive tendency to respond to information in the way outlined above. Confirmation bias can be viewed as one of the effects of a confirmatory value, but the two shouldn't be conflated; social justice, equality, or other values aren't themselves cognitive tendencies. Despite these differences, as their names suggest, confirmation bias and confirmatory values share a crucial functional property: they both drive individuals to predetermined conclusions and impede impartial assessment of the relevant data.

With these points in mind, the argument that I shall develop in the remainder of this article is the following: Research on human reasoning and confirmation bias suggests that because of its functional role, confirmation bias is sometimes Mandevillian in nature, contributing to the reliability of belief-formation at the group level. Since confirmatory values functionally overlap with confirmation bias, they too have that property and thus banning them from science has epistemic costs. It risks weakening the reliability of scientific inquiry. Since it rests on the assumption that confirmatory values always threaten the reliability of science without contributing to it, we should reject the CV view.

The first step in developing this overall argument is to introduce work on human reasoning that suggests that confirmation bias has in some cases, including in scientific inquiries, a Mandevillian profile.

4.1 Recent research on human reasoning and confirmation bias

I will focus in particular on Mercier and Sperber's ([2011], [2017]) work on human reasoning. On the basis of empirical findings and theoretical considerations, Mercier and Sperber argue that, contrary to what is commonly assumed, the evolved function of human reasoning is not so much a means for each individual to discover and track the truth. Rather, human reasoning was selected for argumentative purposes: for (i) convincing other people through argumentation and (ii) evaluating the justifications and arguments other people address to us. This evolutionary thesis is the key component of what Mercier and Sperber ([2011], [2017]) introduce as their 'argumentative theory of reasoning'. It gives rise to a number of predictions. The following two, and the empirical evidence pertaining to them, will be relevant for my discussion below.

Mercier and Sperber hold that if human reasoning evolved to help us convince others, then we should have a confirmation bias when we engage in persuasion. If, say, my goal is to convince you, then I have little use for arguments that support your view or rebut mine; rather, I will benefit from focusing only on information corroborating my point.

Mercier and Sperber ([2011], pp. 63–5) emphasize that the prediction of a confirmation bias in human reasoning is borne out by the data. Many psychologists hold that the bias is 'ubiquitous' (Nickerson [1998]) and 'perhaps the best known and most widely accepted notion of inferential error to come out of the literature on human reasoning' (Evans [1989], p. 41). It is found in everyday and abstract reasoning tasks (Evans [1996]), even if subjects are asked to be more objective (Lord *et al.* [1984]) or paid to reach the correct answer (Johnson-Laird and Byrne [2002]). Its impact also seems to be mostly independent of intelligence and other measures of cognitive ability (Stanovich *et al.* [2013]).

The experimental findings concerning confirmation bias in human reasoning challenge the view that human reasoning has the function of facilitating the acquisition of accurate beliefs in lone thinkers. The bias leads to partial and thus (for the individual) less reliable information processing. The data are, however, exactly as expected if the purpose of human reasoning is to produce arguments that are to persuade others, Mercier and Sperber ([2011], [2017], pp. 206–20) maintain.

Their claim might seem too quick, because if the function of human reasoning is to allow us to better convince others, it should help us to devise strong arguments. Developing strong arguments, in turn, often requires anticipating and addressing counter-arguments. Yet, confirmation bias hinders us in doing just that. It thus seems that if human reasoning evolved to help us better convince others, then *pace* Mercier and Sperber's claim, we should not have such a bias.

Mercier and Sperber ([2017]) respond by noting that anticipating and rebutting objections to one's own view so as to develop compelling arguments takes lone thinkers significant effort and time. Instead, Mercier and Sperber argue, lone thinkers might, and in fact do, adopt a more economical approach. Lone thinkers 'outsource' this cognitive labour by exploiting the interactive nature of dialogue, refining justifications and arguments with the help of the interlocutors' feedback, 'tailoring their arguments to the specific objections raised' (Mercier and Sperber [2017], p. 228). This has the advantage that individual reasoners will only expend as much cognitive effort as is required to persuade others in any given situation (Trouche *et al.* [2016]). And it explains why people are 'lazy' in anticipating objections to their own view and susceptible to confirmation bias, even if the function of human reasoning is to help us better convince others (Trouche *et al.* [2016]).

Turning now to the second prediction of the argumentative theory, if human reasoning evolved so that we are better able to convince others through argumentation and evaluate the arguments that others address to us, as Mercier and Sperber propose, then we should be particularly adept at detecting bad arguments proposed by others. And reasoning should yield superior results in groups than when individuals engage in it alone.

The data support this prediction too, Mercier and Sperber ([2011], [2017]) hold. They review a range of studies suggesting that we are indeed skilled at spotting weaknesses in other people's arguments and even in our own, provided we take the arguments to belong to someone else. For example, Trouche *et al.* ([2016]) asked their test subjects to produce a series of arguments in answer to reasoning problems and afterwards had them quickly assess other people's arguments concerning the same problems. Strikingly, about half of the participants didn't notice that, by the experimenter's slight of hand in some trials, they were presented with their own arguments as if they belonged to someone else. Moreover, among the subjects who accepted the manipulation and thus believed that they were assessing someone else's argument, more than 50% rejected their own arguments. Crucially, they were more likely to do so for invalid than for valid ones. Trouche *et al.* ([2016], p. 2122) thus conclude that people tend to be 'more critical of other people's arguments than of their own'; they are 'better able to tell valid from invalid arguments when the arguments are someone else's than their own'.

These data cohere well with the results of studies involving individual versus group comparisons in reasoning tasks. Studies of this kind found that groups perform better than the average individual, often better than even the best group member (Minson *et al.* [2011]; Maciejovsky *et al.* [2013]). Unsurprisingly, the social exchange of arguments turns out to be critical for improvements in performance (Besedeš *et al.* [2014]; Mellers *et al.* [2014]; Woolley *et al.* [2015]).

Do these considerations hold for the field of science too? Reasoning, understood as the production and evaluation of arguments, is a pervasive process in science. Furthermore, Mercier and Sperber ([2017], pp. 315–17) review experimental (Mahoney [1977]), ethnological (Dunbar [1995]), and historical evidence (Mercier and Heintz [2014]) showing that just like everyone else, scientists are subject to confirmation bias, and better at evaluating other people's arguments than their own. In supporting an extension of the preceding points to scientific reasoning, the data support an account of the latter in which confirmation bias plays a key, Mandevillian role. Building on (Mercier and Sperber [2011], p. 65, [2017], pp. 320-27; Smart [2018], p. 4190), I will now elaborate on that role.

4.2 Mandevillian confirmation bias in science

Consider an example: Suppose there is a group of five scientists trying to answer one of the still open questions in science, such as where life comes from ('primordial soup', a meteorite, and so on). Each of the scientists has a confirmation bias toward a different explanation of the phenomenon. As it happens, none of the five proposals enjoys more empirical success than any other. Suppose the scientists have four weeks to explore the issue and determine the most plausible account among the five views. What would be an epistemically beneficial distribution of research effort within the group? I shall consider two proposals.

Suppose that each of the five scientists can, and is instructed to, impartially assess all five views, and determine the most plausible through group discussion. Suppose too that they all follow the instruction. They suspend their confirmation bias towards their own view and evaluate each of the proposals equally critically and with dispassion.

While this might seem to be the epistemically best distribution of research effort, it has a significant side effect. A confirmation bias towards a particular view, *V*, will tend to push scientists to persistently search for data supporting *V* and to invest effort in defending it. Importantly, in the light of contradictory information that cannot be accommodated by *V*, the bias may incline a scientist to consider rejecting auxiliary assumptions to *V* rather than the proposal itself. In contrast, scientists without the bias are less invested in and committed to *V*, making it more likely that they will engage in a less thorough search for date supporting *V*. Additionally, when encountering information contradicting *V*, or when pressed in group discussions, they may more readily reject the proposal itself, as they simply care less about it. Returning to the example from above, if the five scientists are impartial and unaffected by confirmation bias with respect to all five proposals, there is a risk that each view remains less supported and all theoretical avenues with respect to it less explored than they would be otherwise.

Consider, then, a second way of distributing research effort. Suppose the scientists are allowed to abandon the attempt to even-handedly assess the five proposals and instead are permitted to succumb to their bias towards their own view. Suppose too that, as before, they are instructed to determine the most plausible proposal through group discussion, so that the winning view is the one that survives the most criticism by the most scientists.

In the process of social criticism, their individual confirmation bias will incline each scientist to invest significant effort in gathering data supporting their own view and in responding to counter-evidence and objections in ways that lead to careful exploration and development of the proposal rather than its swift rejection. As a result, since each of the scientists favours one of the five proposals, after four weeks the group will have accumulated more support for the five proposals. And they will have more thoroughly explored them than in the first scenario, putting the group as a whole in an epistemically better position to determine the correct view among the five proposals.

A problem remains: confirmation bias does not reliably track truths (Evans [1989]) and, assuming that only one of the five proposals is correct, then the bias will drive scientists to reach erroneous conclusions. Less invested, less one-sided information processing might thus seem to be more epistemically beneficial for each individual scientist, helping them to avoid exploring misguided proposals.

However, notice that each individual scientist's confirmation bias won't necessarily negatively affect the group's project of determining the most tenable view. Because if, as psychological studies suggest (Trouche *et al.* [2016]), each individual's weakness in critically assessing their own view is offset by a particular strength in detecting flaws in the reasoning of others, then the same should hold for the scientists in the group (Mercier and Sperber [2017], pp. 315–17). As long as the group as a whole pursues the goal of tracking truths and remains flexible,[5] social criticism within the group will help correct, and prompt refinements of, each individual's reasoning, ensuring that the group's conclusions are not too far off target. That is, while confirmation bias may undermine the reliability of belief-formation in each individual, directing most of the five scientists towards mistaken conclusions, the corresponding epistemic risks for the group will be kept in check via social feedback.

Given the specific distribution of epistemic weaknesses and strengths in each individual's reasoning, it now becomes the epistemically most efficient option to distribute research effort in the group so that the five scientists are allowed to give in to their confirmation bias and actively criticize each other's views. This is because if each of the scientists instead suspended their confirmation bias and engaged in impartial information processing, this would result in a more superficial exploration of the hypotheses space. Additionally, the ability to assess the arguments of others is not being effectively exploited if the positions available for criticism are not those that are the most corroborated (*qua* less passionately and thoroughly defended).

Since confirmation bias can thus contribute to the analytical depth of scientific explorations, it can have significant epistemic benefits for scientific groups despite being epistemically detrimental to each individual's reasoning (Mercier and Sperber [2011], [2017]; Smart [2018]). In ensuring a thorough investigation of hypotheses, the bias can increase the reliability of scientific belief-formation and help maximize the acquisition of true beliefs at the group level, provided there is viewpoint diversity and plenty opportunity for social criticism within the group.[6]

4.3 Situating the argument

The argument introduced is related to but also crucially different from a point Solomon ([1992], [2001]) made in an intriguing discussion of case studies from the history of science. Solomon argued that in situations when many theories

[5] This is compatible with most individual scientists being dogmatic. I'll return to the point in Section 5.

[6] There are other conditions that may need to be met; I will return to this point in Section 5.1.

or research programmes enjoy some empirical successes (for example, successful predictions of new phenomena, new explanations of already known phenomena, or successful control and manipulation of processes) but none garners all, it is rational to allocate research effort so that each theory or research programme attains its fair share of attention ([1992], pp. 445–6, [2001], pp. 76–8, 117–19). This will lead to the development of different theories standing in competition with each other, which in turn advances and helps settle scientific debates. Solomon ([1992], pp. 443, 452) maintained that in this situation, cognitive factors such as confirmation bias are epistemically important for groups of scientists, because if each scientist has a confirmation bias toward their own pet theory, this will ensure an equitable distribution of research effort, facilitating the development of and competition between theories.

The argument developed in the previous section coheres well with Solomon's point, but it also differs in two important respects. First, it suggests that confirmation bias is epistemically beneficial not only because it produces a diversity of competing positions, but also because it ensures that these positions and their critiques are more substantially developed than they might otherwise be. Second, Solomon's point that confirmation bias can be epistemically beneficial to science by ensuring a fair distribution of research efforts is relatively weak, given that there are alternative, perhaps less epistemically problematic means to achieve the same ends, such as social systems of reward and sanction (Kitcher [1993]). The argument developed here provides reasons to believe that confirmation bias is likely to be more effective than these alternative means. The bias does the distributional work by harnessing the particular epistemic weaknesses and strengths of each scientist, doing justice to what might well be the evolutionary function of human reasoning (Mercier and Sperber [2017]).

Notice too that alternative mechanisms are likely to rely on the use of money, praise, or other external prompts. These are 'extrinsic' motivations for investing research effort. They are typically contrasted with 'intrinsic' motivations, which are involved when we act without any obvious external rewards (Brown [2007]). Importantly, extrinsic rewards have been found to diminish intrinsic motivation, as subjects tend to interpret them as an attempt to control behaviour (Deci *et al.* [1999]), and studies suggest that extrinsic motivation is frequently less effective than intrinsic motivation (Lepper *et al.* [1973]; Benabou and Tirole [2003]). Ensuring an epistemically beneficial distribution of research resources via these alternative strategies entails that we rely on extrinsic motivations, and this is likely to be less effective than allowing pre-existing, intrinsic motivation (for example, personal or political values) to achieve the same ends.

This completes my argument for the claim that confirmation bias in science has in some cases a Mandevillian character. I shall now relate this to the normative debate on illegitimate values in science.

5 Against the CV View

The CV view rests on the assumption that the functional role of confirmatory values (driving reasoners to predetermined conclusions and hindering an impartial assessment of the data) is epistemically detrimental *per se*, undermining the reliability of scientific inquiry. The preceding discussion of confirmation bias provides reason to question the plausibility of the CV view, suggesting that this functional role can in fact be epistemically beneficial, contributing to the reliability of scientific inquiry at the group level. The CV view appears to be too strong.

It will be useful to illustrate the point by reconsidering the claims of Anderson ([2004]), Douglas ([2009], [2016]), Brown ([2013]), Elliot ([2017]), and Longino ([1990], [2002]). As noted, Anderson ([2004], p. 11) holds that we 'need to ensure that value judgments do not operate to drive inquiry to a predetermined conclusion. This is our fundamental criterion for distinguishing legitimate from illegitimate uses of values in science'.

Considerations laid out here cast doubt on this criterion. Confirmation bias in scientific research can, in some cases, be epistemically beneficial and its suspension epistemically costly, and so we should not attempt to eradicate this bias, and by extension confirmatory values, from science. *Pace* Anderson, attempts to rule out confirmatory values are counterproductive because the standard alternative to confirmatory values—namely, impartiality—is unsatisfactory. Impartiality is likely to result in a more superficial exploration of an impoverished hypothesis space. Anderson's ([2004], p. 11) 'fundamental criterion' for distinguishing legitimate from illegitimate uses of values in science is hence problematic.

The same applies to Douglas's ([2016]) view that values should not be allowed to play a direct role in evidential assessments as this may give rise to wishful thinking. Granted, when confirmatory values affect scientists' reasoning, they may indeed incline scientists to treat evidence that contradicts their favoured hypothesis as less convincing and evidence that supports it as stronger than it is. This does correspond to wishful thinking (Steel [2018]). But these values also equip a scientist with a special sensitivity to a subset of data that more critical researchers might overlook, allowing that scientist to develop a strong case for a favoured conclusion, yielding epistemic benefits at the group level. Thus, that confirmatory values might lead to wishful thinking does not prevent these values playing a legitimate role in science.[7]

[7] In the theorizing on values in science, there has recently been a flurry of research on wishful thinking (de Melo-Martín and Intemann [2016]; Steel [2018]; Hicks and Elliot [unpublished]).

Douglas ([2009]), Brown ([2013]), and Elliot ([2017]) also hold that values are 'unacceptable [in science when they lead inquirers to use] "rigged" methods that generate predetermined outcomes' (Elliot [2017], p. 13), because they will then undermine the 'core value of science—to produce reliable knowledge—which requires the possibility that the evidence produced could come out against one's favoured theory' (Douglas [2009], p. 100). Indeed, when scientists rely on confirmatory values or are affected by confirmation bias, their methods of inquiry are, to some extent, 'rigged' (typically, unconsciously, unintentionally). This follows from aiming to generate support for preferred conclusions. But while Douglas, Brown, and Elliot seem to assume that this threatens reliable belief-formation in science per se, the Mandevillian account of confirmation bias and (by extension) confirmatory values suggests that in some scientific inquiries, the opposite is the case. Confirmatory values and confirmation bias may lead scientists to adopt rigged methods; nonetheless, those values and that bias can contribute to a thorough investigation of a phenomenon at the group level. So even if these values affect an individual scientist such that it is impossible for the evidence collected to 'come out against [the] favoured theory', this does not necessarily undermine the 'core value of science' (to produce reliable knowledge) (Douglas [2009], p. 100). For the evidence could then still come out against their favoured theory at the group level.

Finally, even Longino ([1990], [2002]), who rejects the assumption that objectivity is found in individuals, arguing instead that it is a group-level property, does not acknowledge these group-level benefits of individuals' confirmatory values. Her proposal is to 'limit' the influence of subjective preferences by calling on scientists to 'take up' and respond to critical social feedback and, therewith, contradictory data (Longino [1990], p. 78, [2002], p. 130). Longino's uptake condition is meant to ensure that scientists leave their conclusions open to criticism and revision, rather than anchor their inquiry and response to criticism on a preferred outcome (Biddle [2009]).

But it is important to distinguish between two kinds of uptake, or responsiveness to criticism. There is what I will call 'comprehensive uptake', which involves responding to criticism in ways that leave open the option of abandoning the preferred view. And there is what I will call 'restrictive uptake', which involves responding to criticism in ways that do not leave open this option. Restrictive uptake is clearly required for a group to attain many of the epistemic benefits mentioned in my earlier discussion of the argumentative theory of reasoning. This is because individual's refinement of a favoured position often relies on an 'outsourcing' of cognitive labour (Mercier and

Given the connection between confirmatory values and wishful thinking, the argument developed here offers a contribution to this research.

Sperber [2017], pp. 227–34), and individuals tend to be ‘lazy’ in developing support for their own views until pushed to do so by the objections of others (Trouche *et al.* [2016]). However, comprehensive uptake, which seems to be what Longino calls for, is not required. In fact, since it involves being less committed to one’s favoured view, it is likely to reduce the depth of anaysis developed within scientific groups.

But even when it comes to comprehensive uptake, Longino’s condition does capture an important point. If such uptake never occurred among scientists, the epistemic benefits from confirmatory values could not arise in the group either. For this would preclude the group as a whole from converging on the correct proposal; such convergence presupposes a readiness among the group’s members to update their conclusion(s). However, in order for the group to benefit from confirmatory values, it is not required that each individual exhibit this readiness; it only requires that most of the group, or the group as a whole, do so. If we aim to restrict the influence of confirmatory values so as to ensure that scientific inquiry is as reliable and epistemically efficient as possible, Longino’s proposal of comprehensive uptake is likely to be too strong.

5.1 Qualifications and clarifications

The argument against the CV view rests on an abstract analysis of the potential epistemic benefits of confirmatory values. It sets aside many aspects of the social context in which science actually takes place and assumes scientific environments with (*inter alia*) a diversity of viewpoints, social criticism, and an equal distribution of power and resources among scientists. These conditions are frequently not met in actual scientific research.[8] Since the social conditions in which science takes place play a crucial role in determining whether confirmation bias and confirmatory values are epistemically beneficial, the argument against the CV view needs to be qualified. Specifically, the argument applies only where there exist within the group a diversity of viewpoints, an equal distribution of power, and so on.

This qualification does not undermine the relevance of the argument. It is not implausible to hold that some social environments in science do approach the conditions assumed. Moreover, advocates of the CV view do not limit their concerns to contexts in which these ideal conditions do not prevail. It is an open question whether the CV view is satisfactory in contexts when there is diversity of viewpoint, equality of resource, and so on. The argument developed in this article helps to answer this question.

[8] I’m grateful to an anonymous reviewer for highlighting this and the following points in this section.

A second clarification is required. Particular cases of value-laden research that have worried many philosopher about confirmatory values and biases are cases where, for example, private interests (those of pharmaceutical companies, chemical companies, the fossil fuel industry, and so on) have disproportionate power to fund research and suppress or obscure evidence that would challenge these actors' favoured conclusions (Elliot [2017]). These are cases where the CV view's call for restrictions on the influence of values in science is highly plausible.

Still, the CV view holds that values directing individuals towards predetermined conclusions are epistemically problematic per se and a constraint on them is thus always warranted. The argument offered here is intended to challenge this particular claim only. It is meant to motivate the view that such values are also, in some cases, beneficial and their restriction would be epistemically costly. It may not be easy to strike a balance between allowing confirmatory values to operate and limiting their operation to avoid the pursuit of unpromising avenues or other epistemic costs. But if we treat confirmatory values as always illegitimate in science, as the CV view suggests, then we risk throwing out the baby with the bathwater.

5.2 An objection: The dogmatism problem

The argument against the CV view developed in this article suggests that in some cases confirmation bias and confirmatory values are epistemically beneficial and thus are not illegitimate in science per se. One might object that if we grant that confirmation bias can be acceptable, then we run the risk of allowing dogmatism in science. Because if scientists may ignore evidence and arguments contradicting their favoured conclusions and may limit their search for data to those confirming these conclusions, then they may retain their conclusions in the light of contradictory information and become closed-minded. However, such close-mindedness or dogmatism in science ought to be prevented at all cost. Hence, in line with the CV view, confirmation bias and confirmatory values are likely to be more epistemically pernicious than beneficial, and thus should be considered illegitimate in science. Or so the objection concludes.

Before assessing the point, it is worth clarifying the difference between confirmation bias and dogmatism. As noted, confirmatory bias is the tendency to process information about an issue so that one's pre-existing view about that issue is confirmed, where this also involves ignoring or downplaying contradictory evidence or arguments (Myers and De Wall [2015]). Dogmatism is different. While there are many versions of it, the one I shall focus on here, 'epistemic dogmatism',[9] is commonly taken to be the tendency

[9] Epistemic dogmatism pertains to a scientist's response to, and search for, data within the confines of scientific inquiry. There is also what might be called 'institutional dogmatism', which may involve scientists leaving the scientific field to persuade non-scientific actors, institutions,

to hold a belief 'unquestioningly and with undefended certainty', where this involves a resistance to revising the belief in light of counterevidence (Blackburn [2008], p. 139).

One might be dogmatic in this sense with respect to a certain view without having a confirmation bias related to it. For instance, one might dogmatically hold on to a particular conclusion no matter what data one is presented with and without having the tendency to seek information confirming one's conclusion. Similarly, one might have a confirmation bias with respect to a certain view yet not be dogmatic about the view. For instance, one might tend to confirm one's favoured conclusion and overlook contradictory data, while being open to revising the conclusion when the data are noticed and become strong. Confirmation bias and dogmatism are hence distinct.

They are, however, also closely related. For instance, if one systematically ignores or downplays counterevidence to one's pre-determined conclusion, one will not revise that conclusion in the light of counterevidence. This is a feature of dogmatism (Anderson [2004]). It is the feature of dogmatism to which confirmation bias can clearly contribute and to which the above objection appeals.

The objection would be weak if there was no widespread agreement among philosophers working on values in science that dogmatism about values and viewpoints is indeed generally detrimental to and ought to be prevented in science. But there is. For instance, Longino ([2002]) proposes the 'uptake' condition as a guard against dogmatism in science and often notes that there should be no dogmatism in science (Biddle [2009]).[10] Anderson ([2004]) also insists on the danger of dogmatism. She writes that what is 'worrisome about allowing value judgments to guide scientific inquiry is [. . .] that these judgments might be held dogmatically' ([2004], p. 11). Similarly, Rolin ([2012], p. 211) holds that values are 'an epistemic problem for science insofar as they lead scientists to dogmatism'. In the same vein, Brown ([2013], p. 838) writes that the 'real problem [of values in science is] dogmatism about values'.

No doubt, dogmatism is often problematic in science. But it seems that philosophers who hold that it should always be prevented overlook that dogmatism can also have epistemically beneficial effects in science (Kuhn [1963]). I shall introduce three of them.

advocacy groups, and individuals; to gather funding to support research outside the field of legitimate academic research; to initiate campaigns to promote a favoured view, and so on.

[10] Biddle ([2009]) offers a critique of Longino's proposal that is well in line with my argument. He objects to Longino's assumption that dogmatism is always epistemically problematic: 'Progress in science is best ensured not by demanding of individuals that they be open to everything but, rather, by distributing the resources of a community into various lines of research and letting each of these programs doggedly pursue its own course' (Biddle [2009], p. 622). Biddle does not provide much support for the claim that progress is best ensured if these conditions obtain, however.

Zollman ([2010]) mentions one such benefit. By using a model for network simulation that operates on the basis of a Bayesian update mechanism, Zollman shows that in well-connected networks of undogmatic individuals, false or misleading data can propagate rapidly in the network and is more likely to have a lasting effect on the members' convergence behaviour. In contrast, less well-connected networks, or networks with more dogmatic members, do not face this epistemic risk. Zollman gives a concrete example from the history of medical research on peptic ulcer disease (PUD).

In the 1950s, scientists had to choose between two accounts of PUD. One was the bacterial hypothesis and the other the hyper-acidity hypothesis. The bacterial hypothesis was the correct one and it also enjoyed early evidential support. Yet, in 1954, a prominent gastroenterologist, Eddy D. Palmer, published a study that suggested that bacteria are incapable of colonizing the human stomach (he had looked at more than 1000 patients' biopsies and detected no colonizing bacteria). The result of this study was the widespread abandonment of the bacterial hypothesis in the scientific community. It was not until the 1980s that it became clear that Palmer was wrong. He did not use a silver stain when investigating his biopsies, instead relying on a Gram stain. This matters because *Helicobacter pylori* are most visible with silver stains but are difficult to see with a Gram stain.

Zollman argues that the disproportionate influence of Palmer's publication was partly grounded in a readiness to abandon competing ideas and a lack of dogmatic mindset among advocates of the bacterial hypothesis. This readiness and lack of dogmatism (together with belief perseverance among advocates of the hyper-acidity hypothesis) hindered intellectual progress in the research into PUD for three decades, according to Zollman. He uses this example to illustrate that dogmatism can, in some cases, contribute to the epistemic success of a scientific community by reducing the effect of misleading data, and by sustaining the search for new ideas, methods, and information.

Notice that the nature of the epistemic contribution of dogmatism that Zollman points to is likely to depend on social conditions and power relations. The PUD example, in particular, illustrates that dogmatism pertaining to consensus views (for example, hyper-acidity hypothesis) can be less epistemically beneficial and more problematic than dogmatism pertaining to dissenting views (for example, the bacterial hypothesis).

Turing now to a second positive role that dogmatism might play in science, Popper ([1994], p. 16) notes:

> A limited amount of dogmatism is necessary for progress. Without a serious struggle for survival in which the old theories are tenaciously defended, none of the competing theories can show their mettle—that is, their explanatory power and their truth content.

For Popper, some dogmatism contributes to progress in science, prompting opponents of the dogmatist to make fully explicit, elaborate, and hone their counter-arguments. Indeed, even if the dogmatically held views are entirely misguided, they might still help strengthen and invigorate the deliberative efforts of those who embrace alternatives, stimulating them to make their own proposals more convincing (see also Mill [1998], pp. 22–4, 42–4).

Finally, just as with confirmation bias, dogmatism may benefit science in inclining individuals who encounter strong counterevidence to their pet theory to consider abandoning supplementary hypotheses of the latter when their less dogmatic counterparts would be poised to give up on the entire theory. As a result, there may be situations where dogmatism, like confirmation bias, is crucial in pushing scientists to investigate avenues that would be overlooked by more open-minded individuals (Rowbottom [2011]).

Dogmatism in science is thus not always epistemically problematic. It can provide (i) protection against premature scientific convergence and consensus, (ii) motivation for opponents to better develop their objections and alternative theories, and (iii) a way to ensure all research avenues are explored. These are Mandevillian effects, because at the individual level dogmatism remains epistemically pernicious (reducing one's sensitivity to a subset of data), while at the group level it facilitates (i)–(iii) (Smart [2018]).

Given these points, there is reason to believe that dogmatism in science is not always epistemically bad and should not always be prevented. If this is right, then we will also need to reconsider the objection that confirmation bias and confirmatory values should never be admitted into scientific inquiry because they lead to dogmatism.

6 Conclusion

Dogmatism, confirmation bias, and confirmatory values are perhaps frequently epistemically detrimental in science. The argument of this article was not meant to deny this. The aim was to critically assess the CV view, which says that whenever values drive an individual's and/or a group's inquiry to predetermined conclusions by leading them to skewed, partial processing of information, then these values are epistemically problematic and illegitimate in science. I argued that this view, which many philosophers working on values in science endorse, is too strong. Research on human reasoning and confirmation bias suggests that that bias and, by extension, confirmatory values can have a Mandevillian character in scientific inquiry. That is, despite being epistemically detrimental for individual scientists, in some cases they contribute to the reliability of scientific belief-formation at the group level and facilitate a more in-depth exploration of a given problem space than would otherwise be likely. Since this is so, in treating confirmatory values as illegitimate in science, advocates of the CV view risk undermining their own goal, namely, to support reliable belief-formation and truth-tracking. A more plausible proposal concerning legitimate and illegitimate values in science will need to make room for the possibility that confirmatory values (and dogmatism) can produce epistemic benefits that might, in some cases, make them acceptable parts of science.

Acknowledgements

This article was written while I was a visiting scholar at the University of Cambridge. I'm very grateful to Tim Lewens for many interesting discussions on the issue, and to Edouard Machery and Andreas De Block for helpful feedback on the main argument. Many thanks also to two anonymous reviewers of the journal for comments that helped significantly improved the paper. The research conducted for this paper was funded by the Research Council of KU Leuven (grant agreement no. 3H160214).

Centre for Logic and Philosophy of Science
KU Leuven, Belgium
and
Department of Economics
University College London, UK
uwe.peters@kuleuven.be

References

Alexandrova, A. [2018]: 'Can the Science of Well-Being Be Objective?', *British Journal for the Philosophy of Science*, **69**, pp. 421–45.

Anderson, E. [2004]: 'Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce', *Hypatia*, **19**, pp. 1–24.

Benabou, R. and Tirole, J. [2003]: 'Intrinsic and Extrinsic Motivation', *Review of Economic Studies*, **70**, pp. 489–520.

Besedes, T., Deck, C., Quintanar, S., Sarangi, S. and Shor, M. [2014]: 'Effort and Performance: What Distinguishes Interacting and Non-interacting Groups from Individuals?', *Southern Economic Journal*, **81**, pp. 294–322.

Biddle, J. [2009]: 'Advocates or Unencumbered Selves? On the Role of Political Liberalism in Longino's Contextual Empiricism', *Philosophy of Science*, **76**, pp. 612–23.

Blackburn, S. [2008]: *The Oxford Dictionary of Philosophy*, Oxford: Oxford University Press.

Brown, L. V. [2007]: *Psychology of Motivation*, New York: Nova Publishers.

Brown, M. [2013]: 'Values in Science beyond Underdetermination and Inductive Risk', *Philosophy of Science*, **80**, pp. 829–39.

De Melo-Martin, I. and Intemann, K. [2016]: 'The Risk of Using Inductive Risk to Challenge the Value-Free Ideal', *Philosophy of Science*, **83**, pp. 500–20.

Deci, E. L., Koestner, R. and Ryan, M. R. [1999]: 'A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation', *Psychological Bulletin*, **125**, pp. 627–68.

Douglas, H. [2009]: *Science, Policy, and the Value-Free Ideal*, Pittsburgh, PA: University of Pittsburgh Press.

Douglas, H. [2013]: 'The Value of Cognitive Values', *Philosophy of Science*, **80**, pp. 796–806.

Douglas, H. [2016]: 'Values in Science', in P. Humphreys (*ed.*), *Oxford Handbook of Philosophy of Science*, New York: Oxford University Press, pp. 609–31.

Dunbar, K. [1995]: 'How Scientists Really Reason: Scientific Reasoning in Real-World Laboratories', in R. J. Sternberg and J. E. Davidson (*eds*), *The Nature of Insight*, Cambridge, MA: MIT Press, pp. 365–95.

Elliott, K. [2017]: *A Tapestry of Values: An Introduction to Value in Science*, New York: Oxford University Press.

Evans, J. [1989]: *Bias in Human Reasoning: Causes and Consequences*, Hove: Erlbaum.

Evans, J. [1996]: 'Deciding before You Think: Relevance and Reasoning in the Selection Task', *British Journal of Psychology*, **87**, pp. 223–40.

Fugelsang, J., Stein, C., Green, A. and Dunbar, K. [2004]: 'Theory and Data Interactions of the Scientific Mind: Evidence from the Molecular and the Cognitive Laboratory', *Canadian Journal of Experimental Psychology*, **58**, pp. 86–95.

Haack, S. [2003]: *Defending Science—Within Reason: Between Scientism and Cynicism*, Amherst, NY: Prometheus Books.

Hicks, D. [2014]: 'A New Direction for Science and Values', *Synthese*, **191**, pp. 3271–95.

Hicks, D. and Elliott, K. [unpublished]: 'A Framework for Understanding Wishful Thinking', available at <philsci-archive.pitt.edu/14348/1/Wishful%20Thinking%20final.pdf>.

Intemann, K. [2015]: 'Distinguishing between Legitimate and Illegitimate Values in Climate Modeling', *European Journal of Philosophy of Science*, **5**, pp. 217–32.

Johnson-Laird, P. and Byrne, R. [2002]: 'Conditionals: A Theory of Meaning, Pragmatics, and Inference', *Psychological Review*, **109**, pp. 646–78.

Kahneman, D. [2011]: *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux.

Kitcher, P. [1990]: 'The Division of Cognitive Labor', *Journal of Philosophy*, **87**, pp. 5–22.

Kitcher, P. [1993]: *The Advancement of Science*, Oxford: Oxford University Press.

Kuhn, T. S. [1963]: 'The Function of Dogma in Scientific Research', in A. Crombie (*ed.*), *Scientific Change*, London: Heinemann, pp. 347–69.

Lacey, H. [1997]: 'The Constitutive Values of Science', *Principia*, **1**, pp. 3–40.

Lepper, M., Green, D. and Nisbett, R. [1973]: 'Undermining Children's Interest with Extrinsic Rewards: A Test of the "Overjustification Hypothesis"', *Journal of Personality and Social Psychology*, **28**, pp. 129–37.

Longino, H. [1990]: *Science as Social Knowledge*, Princeton, NJ: Princeton University Press.

Longino, H. [1996]: 'Cognitive and Non-cognitive Values in Science: Rethinking the Dichotomy', in L. Hankinson Nelson and J. Nelson (*eds*), *Feminism, Science, and the Philosophy of Science*, Dordrecht: Kluwer, pp. 39–58.

Longino, H. [2002]: *The Fate of Knowledge*, Princeton, NJ: Princeton University Press.

Lord, C., Lepper, M. and Preston, E. [1984]: 'Considering the Opposite: A Corrective Strategy for Social Judgment', *Journal of Personality and Social Psychology*, **47**, pp. 1231–43.

Maciejovsky, B., Sutter, M., Budescu, D. V. and Bernau, P. [2013]: 'Teams Make You Smarter: How Exposure to Teams Improves Individual Decisions in Probability and Reasoning Task', *Management Science*, **59**, pp. 1255–70.

Mahoney, M. [1977]: 'Publication Prejudices: An Experimental Study of Confirmatory Bias in the Peer Review System', *Cognitive Therapy and Research*, **1**, pp. 161–75.

Mandeville, B. [1705]: *The Grumbling Hive: Or, Knaves Turn'd Honest*, London: Ballard.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K. and Tetlock, P. [2014]: 'Psychological Strategies for Winning a Geopolitical Forecasting Tournament', *Psychological Science*, **25**, pp. 1106–15.

Mercier, H. and Sperber, D. [2011]: 'Why Do Humans Reason? Arguments for an Argumentative Theory', *Behavioral and Brain Sciences*, **34**, pp. 57–111.

Mercier, H. and Heintz, C. [2014]: 'Scientists' Argumentative Reasoning', *Topoi*, **33**, pp. 513–24.

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

Mill, J. S. [1998]: *On Liberty*, Philadelphia, PA: Pennsylvania State University Press.

Minson, J. A., Liberman, V. and Ross, L. [2011]: 'Two to Tango', *Personality and Social Psychology Bulletin*, **37**, pp. 1325–38.

Morton, A. [2014]: 'Shared Knowledge from Individual Vice: The Role of Unworthy Epistemic Emotions', *Philosophical Inquiries*, **2**, pp. 163–72.

Myers, D. and DeWall, N. [2015]: *Psychology*, New York: Worth Publishers.

Nickerson, R. [1998]: 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises', *Review of General Psychology*, **2**, pp. 175–220.

Peters, U. [2016]: 'Human Thinking, Shared Intentionality, and Egocentric Biases', *Biology and Philosophy*, **31**, pp. 299–312.

Peters, U. [forthcoming]: 'Implicit Bias, Ideological Bias, and Epistemic Risks in Philosophy', *Mind and Language*, available at <doi.org/10.1111/mila.12194>.

Popper, K. [1994]: 'The Myth of the Framework', in M. A. Notturno (*ed.*), *Defence of Science and Rationality*, London: Routledge.

Rolin, K. [2012]: 'Feminist Philosophy of Economics', in U. Mäki (*ed.*), *Philosophy of Economics*, Amsterdam: Elsevier, pp. 199–217.

Rooney, P. [1992]: 'On Values in Science: Is the Epistemic/Non-epistemic Distinction Useful?', *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, **1992**, pp. 13–22.

Rowbottom, D. [2011]: 'Kuhn vs. Popper on Criticism and Dogmatism in Science: A Resolution at the Group Level', *Studies in History and Philosophy of Science*, **42**, pp. 117–24.

Smart, P. [2018]: 'Mandevillian Intellingence', *Synthese*, **195**, pp. 4169–200.

Solomon, M. [1992]: 'Scientific Rationality and Human Reasoning', *Philosophy of Science*, **59**, pp. 439–55.

Solomon, M. [2001]: *Social Empiricism*, Cambridge, MA: MIT Press.

Stanovich, K., West, R. and Toplak, M. [2013]: 'Myside Bias, Rational Thinking, and Intelligence', *Current Directions in Psychological Science*, **22**, pp. 259–64.

Steel, D. [2018]: 'Wishful Thinking and Values in Science', *Philosophy of Science*, **85**, pp. 895–905.

Trouche, E., Johansson, P., Hall, L. and Mercier, H. [2016]: 'The Selective Laziness of Reasoning', *Cognitive Science*, **40**, pp. 2122–36.

Wilholt, T. [2009]: 'Bias and Values in Scientific Research', *Studies in History and Philosophy of Science*, **40**, pp. 92–101.

Woolley, A. W., Aggarwal, I. and Malone, T. W. [2015]: 'Collective Intelligence and Group Performance', *Current Directions in Psychological Science*, **24**, pp. 420–4.

Zollman, K. [2010]: 'The Epistemic Benefit of Transient Diversity', *Erkenntnis*, **72**, pp. 17–35.