CrossMark

episteme

ARTICLE

Can We Talk It Out?

Miguel Egler [iD]

Tilburg University, TiLPS, the Netherlands
Email: m.egler@uvt.nl

Abstract

Research on the normative ideal of democracy has taken a sharp deliberative and epistemic turn. It is now increasingly common for claims about the putative *cognitive* benefits of political deliberation to play central roles in normative arguments for democracy. In this paper, I argue that the most prominent epistemic defences of deliberative democracy fail. Relying on empirical findings on the workings of implicit bias, I show that they overstate the epistemic virtues of political deliberation. I also argue that findings in cognitive and social psychology can aid in the development of a new and improved generation of epistemic arguments for deliberative democracy.

Keywords: epistemic democracy; deliberative democracy; political deliberation; implicit bias; stereotypes

1. Introduction

Over the last 30 years, research on the normative ideal of democracy has taken a decisive deliberative turn. It is now commonplace to regard political deliberation as an essential feature of the democratic political decision-making process (Bächtiger *et al.* 2018). Proponents of this view have traditionally sought to justify their claims by appealing to the morally desirable features of political deliberation. For example, Gutmann and Thompson (2009) argue that procedures of political deliberation make political decisions legitimate because they help safeguard the principles of reciprocity and the dignity of citizens in the democratic process. And both Cohen (2003) and Benhabib (1996) maintain that decisions reached via procedures of political deliberation are inherently just. Going against the grain of these traditional approaches, there has been growing interest in the idea that deliberative democracy is justified (at least in part) by its putative *epistemic* virtues (Estlund and Landemore 2018). In this paper, I argue that the most prominent epistemic arguments for deliberative democracy to date fail. Using empirical findings on the workings of implicit bias, I show that all such arguments rely on exaggerated claims about the epistemic benefits of political deliberation. However, I also argue that research on the workings of implicit biases offers valuable resources to help develop a new and improved generation of epistemic arguments for deliberative democracy that can sidestep these shortcomings.

I first provide an overview of the most prominent epistemic arguments for deliberative democracy in the recent literature (section 2). I then argue that these epistemic arguments are unwarranted insofar as they overstate the epistemic benefits of political deliberation (section 3). My argument is two-fold. First, I explain that epistemic arguments for deliberative democracy are conditional on deliberation being more likely to correct for implicit biases than other relevant methods of political decision-making. Second, I argue that empirical findings suggest that deliberation will systematically fail to correct for a whole class of implicit biases. Taken together, this suggests that for a host of cases, deliberation is no more likely to correct for biases than alternative political decision-making procedures. Lastly, I explain how these findings can help us make better-informed assessments of when we can expect political deliberation to indeed improve the epistemic quality of political decisions (section 4).

Before proceeding, two important clarifications. First, there is an ongoing (and thorny) dispute on what precisely counts as deliberation under the normative ideal of deliberative democracy. Deliberation is typically defined as a mode of *interpersonal* communication geared towards the exchange of reasons and arguments about political views with the aim of solving issues of common concern (Mansbridge *et al.* 2010). However, many have criticised this definition on the grounds that it excludes more 'passion-driven' or 'emotionally-rooted' forms of expression – which, arguably, *should* have a role in political decision-making (see, e.g., Young 1996). In light of these and other similar kinds of critiques, there is now considerable support for the idea that we should turn to a more inclusive notion of what counts as deliberation. For what it's worth, I am highly sympathetic to these critiques. However, most epistemic arguments for deliberative democracy operate with the (more restrictive) traditional picture of deliberation. In order to dialogue better with these arguments, which are the target of this paper, I will operate with this conception here as well.

Second, I use the term 'implicit bias' to refer to the inadvertent reliance on stereotypes and prejudices in social cognition. There has recently been significant scepticism about research on implicit bias, with concerns ranging from minor complaints about the methodology of some studies to more forceful and all-encompassing suspicions about the validity of social implicit cognition research overall.[1] In light of such growing criticism, some argue that philosophers should refrain from relying on the current body of work on implicit bias – at least until such concerns have been adequately addressed.[2] However, some have taken pains to show that although there is indeed room for improvement in research on implicit bias, the all-encompassing critiques of the field turn out to be unwarranted. I cannot, of course, resolve such thorny and intricate disputes in this paper. Rather, I will simply take the arguments here as conditional on work on implicit bias not being fully discredited. However, throughout the paper, I indicate points of contention in the empirical literature I rely on.

2. Epistemic Arguments for Deliberative Democracy

Epistemic arguments for deliberative democracy can be divided into two broad camps: instrumental and non-instrumental (Peter 2016). In this section, I provide an overview

---

[1]For meta-analyses of findings, see Oswald *et al.* (2013) and Forscher *et al.* (2019). For critiques, see Machery (2016), Hermanson (2017) and Buckwalter (2019).

[2]For instance, Machery (2017) claims it would be 'irresponsible' to theorise on the basis of work on implicit biases.

of the most prominent formulations of each one. This will prove important for the arguments I develop in the following sections.

2.1. Instrumental Epistemic Arguments

Instrumental epistemic arguments have two core claims:[3]

    (i) Political deliberation is more likely to produce correct political decisions than relevant alternative methods of political decision-making.

   (ii) Deliberative democracy is justified (at least in part) because political deliberation is more likely to produce correct political decisions than relevant alternative methods of political decision-making.

Two important clarifications are in order. First, there are different views about what counts as a *relevant* alternative political decision-making method. Some take the relevant alternatives to be restricted to democratic methods (e.g., Estlund 2008), whereas others regard *any* such methods to be relevant (e.g., Landemore 2017). We need not adjudicate between these two approaches here. For current purposes, we should only note that all parties to these debates regard (democratic) majority voting to be a relevant alternative method to political deliberation.

Second, this instrumental argument can be fleshed out in different ways, depending on what counts as a *correct* political decision. Notably, many reject that we can evaluate political decisions as such (Rawls 2005; Urbinati 2014; Schwartzberg 2015). Objections span from worries about the theoretical usefulness of these evaluations to suspicions about their very intelligibility. I do not know of a knock-down argument against these different objections. But so as to not load the dice against instrumental epistemic arguments from the start, I set this critique aside. My aim in what follows will be to concede as much as possible to proponents of instrumental epistemic arguments to assess such proposals by their own lights. For this aim it is worth asking: what would it mean to say that the decision is correct in a sense that is *epistemically* significant? Below I list some relevant options.

Truth A political decision is *correct* only if it is produced by a process that tracks the truth.[4]

Knowledge A political decision is *correct* only if it is based on[5] knowledge.[6]

Justified A political decision is *correct* only if it is based on beliefs that are (epistemically) justified (or rational).[7]

Error-avoidance A political decision is correct only if it avoids errors.[8]

---

[3]For (incomplete) surveys of these arguments, see Martí (2006) and Estlund and Landemore (2018).

[4]For reviews of such views, see Estlund and Landemore (2018), Goodin and Spiekermann (2018) and Misak (2002).

[5]The notion of 'basing' in epistemology is controversial. I remain neutral on what this amounts to.

[6]See, e.g., Landemore (2017: Ch. 8).

[7]A notable defender of this view is Gaus (1996).

[8]See Bohman (2007).

There are of course other possible options. For instance, one might argue that political decisions are correct only if stemming from an *understanding* of the relevant political issues. Alternatively, we could say that political decisions are correct only if they accord with precepts of epistemic utility theory.[9] And although all these options are plausible, there has not yet been any attempt to develop *instrumental* epistemic arguments along these lines. Since my aim is to focus on what are currently the most prominent formulations of these arguments in the literature, I will set aside these alternative (unexplored) options.

The above interpretations of what counts as a correct political decision deliver the following distinct readings of the first core claim of instrumental epistemic arguments for deliberative democracy:

> **i-Truth** Political deliberation is more likely to produce political decisions that *track the truth* than relevant alternative methods of political decision-making.

> **i-Knowledge** Political deliberation is more likely to produce political decisions that are *based on knowledge* than relevant alternative methods of political decision-making.

> **i-Justified** Political deliberation is more likely to produce political decisions that are *based on epistemically justified beliefs* than relevant alternative methods of political decision-making.

> **i-Error-avoidance** Political deliberation is more likely to produce political decisions that *avoid (epistemically) bad outcomes* than relevant alternative methods of political decision-making.

Each one of the above readings motivates a version of an instrumental epistemic argument for deliberative democracy, as each one can be used to support (ii) – i.e., the conclusion that deliberative democracy is partly justified by its tendency to produce correct political decisions. However, these distinct formulations are not mutually exclusive. Indeed, some are arguably closely linked. For instance, it is almost universally accepted that knowledge is factive. Moreover, many also accept that having knowledge that p entails having a justified belief that p.[10] This suggests that defences of i-Knowledge require at least some version of i-Truth and i-Justified. And since many of the prominent defences of instrumental epistemic arguments for deliberative democracy are committed to some version of i-Knowledge (Nino 1996; Estlund 2008; Landemore 2017; cf. Peter 2009), then we should plausibly read them as committed to i-Truth and i-Justified as well. Likewise, it is plausible to read defences of i-Truth as defences of i-Error-avoidance. After all, political decisions that result from truth-tracking processes will presumably help avoid (epistemically) bad outcomes. Nevertheless, some of these claims can be arguably defended individually as well. For instance, Gaus (1996) is plausibly read as a proponent of i-Justified, but not i-Knowledge or

---

[9]For a discussion of epistemic utility theory, see Pettigrew (2016).

[10]On the traditional view, justified belief is a necessary condition for knowledge; however, in recent years some have defended that knowledge is what gives rise to justified belief. For a discussion, see Ichikawa and Steup (2018).

i-Truth.[11] And Bohman ([2007](#)) defends i-Error-avoidance while abstaining from i-Truth.

Furthermore, all the above readings of the first premise of instrumental epistemic can *themselves* be interpreted in different ways. For instance, interpretations of i-Truth will vary depending on which conception of *truth* we adopt. On a correspondence view of truth, a political decision is the product of a truth-tracking process only if it is likely to *correspond* to a particular set of facts (moral, meta-ethical, modal, or facts of another sort). However, we can also read Truth in terms of a pragmatic conception of truth. One plausible way of developing this proposal states that a decision is the product of a truth-tracking process only if it is likely to not be falsified by our best methods of inquiry.[12] We can also interpret Truth within a particular constructivist notion of truth – whereby a decision is the product of truth-tracking processes only if members of a collective would make that decision when engaged in an *ideally* rational process of deliberation. This list is not exhaustive.[13] But this will not matter for our purposes. Instead, what is important to emphasise is that under all of these readings, political decisions are correct only if they accord with *some* relevant aspect of the world – whether facts, outcomes of ideal deliberation, methods of inquiry, or something else.[14]

Let's now consider i-Knowledge. The particular notion of truth we adopt will of course matter for how we interpret this claim. Knowledge is widely regarded to entail truth. Thus, different readings of truth will deliver distinct verdicts on what one knows or does not know. For instance, an account of what it means to say a political decision is based on knowledge will vary depending on whether we adopt a pragmatist or correspondence notion of truth.

Turning now to i-Justified. Our interpretation of this proposal will vary depending on which conception of epistemic justification we adopt. The most notable distinction among such views is between internalist and externalist views of justification (see, e.g., Alston [1989](#)). On extreme internalist views, a political decision is correct only if it is warranted by mental states accessible to the decision-makers. By contrast, views of epistemic justification in the externalist extreme of the spectrum will regard political decisions as correct only if the decision-making process used is in fact reliable. There are many further distinctions among views of epistemic justification – e.g., disputes on whether justification requires safety, or whether it's probabilistic. These further distinctions will not prove important here. Rather, it's crucial to note that all views of epistemic justification lie somewhere along the spectrum of the internalist/externalist divide.

Lastly, let's examine i-Error-avoidance. Recall that on this proposal, decisions are correct only if they avoid (epistemically) bad outcomes. A plausible reading of this claim has it that political deliberation prevents decisions based on misinformation or on mistaken reasoning. But it can also mean (more narrowly) that political deliberation prevents decisions that are not influenced by the deleterious effects of biases (see, e.g. Bohman [2007](#)).

---

[11]Arguably, we can also read Habermas ([1996](#)) in this same way.

[12]For an elaboration of this view, see Misak ([2002](#)). Arguably, Anderson ([2006](#)) also defends a version of this view.

[13]Other relevant options include deflationist views of truth (e.g., Horwich [1998](#)) and Cohen's ([2009](#)) *political* notion of truth.

[14]I'll refrain from adjudicating at this point among these different readings as my argument will apply equally to all of them.

In sum, there are many ways of fleshing out instrumental epistemic arguments, depending on what counts as a correct political decision and on how we interpret the relevant epistemic notion in these different readings. The interpretations discussed here are not exhaustive; however, they represent the most prominent approaches to fleshing out this kind of argument in the recent literature. Let's now turn to another kind of epistemic argument for deliberative democracy – one that does not depend on the notion of a *correct* political decision.

2.2. Non-instrumental Epistemic Arguments

Non-instrumental epistemic arguments claim that deliberation *itself* has epistemic value. In this sense, they contend that political deliberation is more likely to promote *responsible* epistemic agency than alternative methods of political decision-making. Relying on this claim and on the assumption that we should try to be responsible epistemic agents, proponents of non-instrumental epistemic arguments defend deliberative democracy. More schematically:

(i) Political deliberation is more likely to promote responsible epistemic agency than relevant alternative methods of political decision-making.
(ii) Deliberative democracy is justified (at least in part) because political deliberation is more likely to promote responsible epistemic agency than relevant alternative methods of political decision-making.

Three clarifications of this argument are in order. First, we can again set aside debates about what counts as a relevant alternative political decision-making method. For current purposes, we have only to keep in mind that (democratic) majority voting is widely regarded to be one such relevant alternative method. Second, note that a major upshot of non-instrumental epistemic arguments is that the epistemic benefits of deliberative democracy are not *conditional* on its potential to promote correct decisions. However, non-instrumental epistemic arguments are not incompatible with this claim either. Lastly, there are many ways of fleshing the notion of responsible epistemic agency at issue in these arguments. I focus on two readings in particular, which correspond to the main non-instrumental epistemic arguments in the literature to date. The first is due to Robert Talisse, and the second to Fabienne Peter.

Talisse's (2009) non-instrumental epistemic argument for deliberative democracy relies on three main ideas. The first is that deliberation helps people acquire more evidence about distinct positions and views about political matters. The second is that political beliefs based on evidence are epistemically superior to those beliefs which are not. The third is that if people hold political beliefs, they should aim to have *true* political beliefs. This means that agents should aim to be responsible epistemic agents with regards to their political beliefs – and that as such, they should *at least* aim to have beliefs that are adequately responsive to evidence on political issues. From the first and second claims, Talisse (2009) argues that political deliberation *itself* has epistemic value as it favours the formation of political beliefs based on evidence. And from the third claim, Talisse contends that people should endorse some form of deliberative democracy. After all, as deliberation facilitates the formation of political beliefs that align with the evidence, and if people should aim to be epistemically responsible agents with regards to their political beliefs, then they should favour a model of deliberative democracy. We then get the following argument:

i-Evidence Political deliberation is more likely to promote the formation of political beliefs based on evidence than relevant alternative methods of political decision-making.

ii Deliberative democracy is justified (at least in part) because political deliberation is more likely to promote responsible epistemic agency than relevant alternative methods of political decision-making.

Peter (2013) offers an alternative non-instrumental epistemic argument for deliberative democracy. Like Talisse, she claims that the procedure of deliberation promotes responsible epistemic agency and that people should aim to be epistemically responsible with regards to their political beliefs. However, Peter takes on another tack to defend the epistemic value of deliberation. Her main contention is that debates about the epistemology of peer disagreement prove relevant to flesh out this idea. She focuses on two central claims. First, she contends that some instances of peer disagreement are persistent: parties to such disputes will remain divided, even when they have adequately considered all the available evidence on the matter. Second, she maintains that in at least some cases of persistent peer disagreement, parties to the dispute should adjust their beliefs to accommodate for the disagreement. Building on these two claims, she argues further that the epistemic value of political deliberation rests in its potential to uncover such cases of persistent peer disagreement in a democratic collective and to make agents aware that they need to adjust their beliefs so as to weigh in disagreement on matters of public concern. Peter thus contends that political deliberation enables members of a collective to become epistemically responsible agents insofar as it enables them to remain (epistemically) mutually accountable to each other. Moreover, Peter contends that since it is clear we should be epistemically responsible agents, then we should prefer models of public governance that can make us aware of potential disagreements. This then yields the following argument:

i-Accountability Political deliberation is more likely to promote mutual epistemic accountability than relevant alternative methods of political decision-making.

ii Deliberative democracy is justified (at least in part) because political deliberation is more likely to promote responsible epistemic agency than relevant alternative methods of political decision-making.

Thus, there are different ways of developing a non-instrumental epistemic argument for deliberative democracy. These different readings will vary in accordance with how we flesh out the notion of responsible epistemic agency. In this section, I have focused on two readings, which map onto the two most prominent non-instrumental epistemic arguments advanced so far in the literature.

3. On implicit biases
In this section, I argue that the most prominent epistemic arguments for deliberative democracy discussed in the previous section are unwarranted because they exaggerate the epistemic benefits of political deliberation. More schematically, I argue that:

P1 The most prominent epistemic arguments for deliberative democracy are conditional on the potential of deliberation to mitigate implicit biases.

P2 If it is likely that deliberation systematically fails to mitigate whole classes of implicit biases, then the most prominent epistemic arguments for democracy are unwarranted.

P3 It is very likely that deliberation systematically fails to mitigate whole classes of implicit biases.

C The most prominent formulations of epistemic arguments for deliberative democracy are unwarranted.

I take it that the inference in P2 is self-evident: given P1, it is clear that if deliberation most likely fails to mitigate a whole class of implicit biases, then the epistemic arguments outlined in previous sections are unwarranted. In what follows, I take P2 for granted and focus on developing and defending P1 and P3.

3.1. P1

As a first step, it is useful to emphasise that P1 is not an empirical claim. It simply states a condition for (instrumental and non-instrumental) epistemic arguments for democracy to go through. To defend this premise, it will then suffice to focus on a *fictional* scenario:

> *Quota*: A *cognitively* diverse collective C deliberates on whether recent economic indicators warrant changes in the country's current immigration quota. At first glance, the data seem mixed. But careful analysis of the evidence reveals that immigrants have an overall positive impact on the country's economy – and that it would be best to either keep quotas as they are, or to increase them. But the majority of the members of C are inadvertently prone to implicit biases against immigrants, which causes them to misinterpret the relevant evidence – *despite their explicit beliefs being favourable towards immigrants*. The collective C eventually decides (by a tight margin) to reduce the immigration quota. But if all members of C had adequately considered the evidence, they would have kept the quotas at their current levels.

Two important clarifications about *Quota* are in order. First, as stipulated, the collective C is *cognitively* diverse. Following Landemore (2017: Ch. 4), I take this to mean that due to differences in perspectives, interpretations, heuristics and predictive models, members of C adopt distinct cognitive approaches in problem-solving. Moreover, I grant the influential idea that a *cognitively* diverse group will more likely arrive at correct political decisions when deliberating together (Landemore 2017). To a first approximation, this might seem to raise a problem: how can members of a cognitively diverse group have the *same* implicit bias against immigrants? However, we can modify this scenario so that members of C have a variety of *different* kinds of implicit biases against immigrants – all of which make them prone to misinterpret the relevant evidence.

I now turn to instrumental epistemic arguments and argue that they are all conditional on the potential of deliberation to correct for implicit biases like the one described in *Quota*. My argument proceeds in two steps. I begin by arguing that under all readings of instrumental arguments discussed above, the decision reached by collective C turns out to be *incorrect*. I then explain why this reveals that

instrumental epistemic arguments are thereby conditional on the potential of deliberation to mitigate implicit biases.

Let's begin with instrumental epistemic arguments that rely on Truth. These arguments state that political deliberation tends to yield correct political decisions because these decisions are produced by a *truth-tracking* process. As we saw, this proposal can be fleshed out in different ways, depending on whether we adopt, say, a correspondence, pragmatist, or constructivist notion of truth. I contend that under all such readings, the decision reached by C is not correct. After all, their decision does not *correspond* with the fact that immigrants have a positive impact on the economy. And as stipulated above, members of C would have chosen differently if they had adequately considered this evidence. Their decisions are then, first, not consistent with what would be the result of adequate inquiry nor, second, the result of entirely rational deliberation on this matter. So, the decision in *Quota* is not the product of a truth-tracking process – whether we adopt a correspondence, pragmatist or constructivist view of truth. As already mentioned, these readings of Truth are not exhaustive (for instance, we have not considered deflationist or political notions of truth). However, I maintain that minor modifications to *Quota* can easily accommodate for these different accounts and yield the same result: that is, the deleterious effect of the implicit bias against immigrants will lead to a decision that does not accord with whatever aspect of the world is relevant for settling what counts as *truth*.

Furthermore, the decision reached by C will thereby also be incorrect according to instrumental epistemic arguments based on Knowledge. For, given that the decision to reduce immigration quotas is likely not likely to be true in any of the above relevant interpretations, then this implies that this decision is thereby not based on knowledge.

Let's now consider instrumental epistemic arguments that focus on Justification. As discussed, their first premise states that political deliberation is more likely to produce political decisions that are *based on epistemically justified beliefs* than relevant alternative methods of political decision-making. Moreover, we saw that our interpretation of this premise varies depending on where our adopted notion of justification lies along the internalist/externalist spectrum. Critically, the decision reached by C in *Quota* is plausibly incorrect on all such readings. If the implicit bias afflicting members of C causes them to engage in stereotype-based reasoning that leads them to inadequately dismiss relevant evidence about economic indicators, then their decision will thereby not stem from the operation of a reliable process.[15] Moreover, their decision will also not be warranted by mental states accessible to those members of C. As stipulated above, members of C are aware of all the available evidence on the matter – and so, even on the extreme internalist/externalist views of epistemic justification, the decision reached by C is epistemically unjustified.

Lastly, consider again instrumental epistemic arguments that turn on Error-avoidance. Recall that the first premise of these arguments is that political deliberation is more likely to produce political decisions that *avoid (epistemically) bad*

---

[15]Stereotype-based reasoning is generally reliable. This may raise doubts that relying on stereotypes about immigrants to make political decisions in *Quota* is unreliable. In effect, this doubt is connected the well-known generality problem for reliabilist epistemology: how do we define which specific cognitive process is causally responsible for a token doxastic state. On one solution to this problem, the fact that C's decision in *Quota* is the result of stereotype-based reasoning makes it reliable. I cannot of course solve the generality problem here. But this particular solution seems wrong. After all, it would give carte blanche to relying on any stereotype whatsoever in reasoning – no matter how racist or mistaken.

outcomes. The most prominent defence of this view states that political deliberation leads to correct decisions by diminishing the chance that cognitive biases deleteriously influence political decisions (Bohman 1998). On this definition, the implicit bias described in *Quota* leads to an epistemically bad outcome. After all, the bias hampers the abilities of members of C to adequately examine the available evidence – and so, their final decision is thereby epistemically problematic. So, on the most prominent reading of Error-avoidance, the decision reached by C in *Quota* cannot be correct.

In sum, these considerations suggest that under all relevant readings of instrumental epistemic arguments, members of C reached an incorrect political decision on matters of immigration quotas. Moreover, they reached an incorrect decision despite having deliberated on the matter. This means that in the case of *Quota*, political deliberation failed to promote a political decision that was likely to be true, based on knowledge or justified belief, or to lead members of C to make decisions that avoid (epistemically) bad outcomes – such as basing decisions on misinformation and mistaken reasoning.

Crucially, these considerations suggest that in *Quota*, political deliberation turns out to be *no more likely to produce decisions that are the result of truth-tracking processes than democratic majority voting*. After all, if deliberation proves ineffective at correcting for the implicit bias against immigrants, then presumably members of C would issue the same verdict on matters of immigration quota if they were simply asked to vote independently. Thus, the case of *Quota* is *one* case where political deliberation is no more likely to produce a correct political decision when compared with a relevant alternative method of political decision-making – viz., majority voting. In other words, this reveals how the instrumental epistemic arguments analysed above are conditional on the potential of deliberation to mitigate implicit biases.

This single instance in which political deliberation fails to give rise to a correct political decision is, of course, not *itself* troubling. The first claim of instrumental epistemic arguments merely states that political deliberation is *more likely* to give rise to correct political decisions when compared with alternative decision-making methods – not that it *always* does. So this isolated instance described in *Quota* does not undermine such arguments. However, it is noteworthy that beyond the implicit bias described in *Quota*, there are other kinds of implicit biases that have potentially deleterious effects on decision-making.[16] And if these implicit biases are also as pervasive as the one described in *Quota*, then they would presumably hamper correct political decision-making in similar ways. Thus, the failure of political deliberation in *Quota* shows how deliberation may systematically fail to produce correct political decisions in a host of other instances as well. And if this is indeed the case, then deliberation will be no more likely to correct for such implicit biases than democratic majority voting. Thus, cases of implicit biases cast doubt on the first core claim of instrumental epistemic arguments for deliberative democracy.

Turning now to non-instrumental epistemic arguments. Recall, the first premise of these arguments states that political deliberation is more likely to promote responsible epistemic agency than relevant alternative methods of political decision-making. The scenario described in *Quota* also proves relevant for an assessment of these arguments. After all, this bias presumably hampers responsible epistemic agency in the sense that proves relevant for these arguments. To see why, note that the bias in *Quota* shows that members of C will fail to form political beliefs that are based on the evidence on matters of immigration. Indeed, as described in the scenario, the deleterious effect of this bias

---

[16]For a review, see Brownstein and Saul (2016).

makes it that members of C will continue to misinterpret the relevant evidence on the matter despite deliberating on it. Thus, deliberation is no more likely to foster the formation of beliefs based on evidence than other methods of political decision-making such as majority voting. Moreover, it is also clear that this bias prevents responsible epistemic agency in the sense of ensuring mutual accountability between members of C. After all, the bias in question prevents certain members of C from accommodating the views of others who disagree with them on these matters. As such, this bias against immigrants will make it unlikely that members of C will adjust their beliefs in response to disagreements that turn on considering the evidence in question. And to the extent that this failure to promote responsible epistemic agency replicates for other cases of implicit biases, then the case described in *Quota* is not an isolated case. Rather, it illustrates a systematic pattern that is likely to reproduce for other kinds of implicit biases as well. So cases of implicit biases putatively pose a challenge for the first premise of non-instrumental epistemic arguments for deliberative democracy.

These considerations demonstrate that epistemic arguments for deliberative democracy are conditional on the potential of deliberation to mitigate the deleterious effects of implicit biases. This is a requirement for instrumental epistemic arguments or else they fail to establish that deliberation is more likely to produce correct political decisions when compared with relevant alternative methods of political decision-making (e.g., democratic majority voting). And likewise, this mitigating effect is a requirement for non-instrumental epistemic arguments, or else they fail to establish that deliberation promotes responsible epistemic agency on the part of members of a deliberating collective. This then raises the question: to what extent does political deliberation prove apt to mitigate the deleterious effects of implicit biases? I address this question in the next section.

3.2. P3

In this section, I argue for P3 – that is, the claim that deliberation will most likely systematically fail to mitigate a host of implicit biases. To begin, it is important to emphasise that there is a dearth of *direct* empirical evidence on the effects of collective deliberation on people's implicit attitudes. Nevertheless, it is useful to begin by reviewing a few findings that suggest deliberation may indeed help mitigate *some* implicit biases.[17]

Recall that deliberation is a mode of mutual communication geared towards the exchange of arguments and reasons for political positions and views. Moreover, during the course of deliberation, we can expect that some people may end up being persuaded by arguments for political positions that are favourable towards certain marginalised social groups. And, as some evidence suggests, the logical force of arguments favourable to social groups can prove effective at mitigating implicit biases against them (Brinol *et al.* 2009). In addition, it is reasonable to expect that in deliberation, people may be led to consider examples and scenarios that contradict their negative stereotypes and prejudices about such groups. And as empirical findings show, this is a technique which has also proven effective at mitigating implicit biases (Blair *et al.* 2001). Furthermore, deliberation also provides the conditions for people to take on different perspectives and focus on individual (rather than stereotypical) features of members

---

[17]For a thorough discussion of work on debiasing, as well as replies to critiques to this literature, see Madva (2017).

of social groups. This is yet another technique which can correct for biases (Galinsky and Moskowitz 2000).

In sum, the evidence seems to favour thinking that deliberation has considerable debiasing potential. However, developments from two active lines of research on implicit biases challenge this thought. The first comes from inquiry into their *nature*. In recent years, there has been a consensus forming that the processes subsumed under the category 'implicit biases' do not form a homogeneous bunch. This claim has at least some degree of initial plausibility. Even on the surface, implicit biases against transgender people appear significantly different from implicit biases against certain political ideologies. This suggests there is reason to believe these two types of biases are underwritten by distinct kinds of psychological mechanisms. This idea gains further traction in light of a host of empirical findings. Most notably, people's scores on tests measuring for implicit biases often do not correlate (for a review: Nosek *et al.* 2007).[18] For instance, a person who scores highly on tests measuring for an implicit bias that associates black people and negative affect can nevertheless get a low score in tests measuring for an implicit bias associating black people to physical characteristics (e.g., being athletic and rhythmic) – and vice versa. Since people's scores are thus disassociated, then there is good reason to think that the implicit bias tested in these studies must be produced by distinct mechanisms. In this sense, many have suggested that rather than speaking of implicit biases as a unified category, we would do best to group implicit biases according to the kinds of psychological processes giving rise to them, as this would better reflect the empirical reality (Holroyd and Sweetman 2016; Del Pinal *et al.* 2017; Del Pinal and Spaulding 2018; Madva and Brownstein 2018).

The second comes from research on debiasing techniques – that is, interventions aimed at correcting for the deleterious effects of implicit biases.[19] As many have pointed out, normative recommendations for combating a given implicit bias must be sensitive to the mechanisms giving rise to it (Holroyd and Sweetman 2016; Del Pinal and Spaulding 2018; Madva and Brownstein 2018). For instance, consider once more the evidence discussed above that implicit biases associating black people to physical characteristics are underwritten by distinct mechanisms from those of the implicit bias associating black people to negative affect. As Holroyd and Sweetman (2016) point out, these findings indicate that a debiasing technique that has proven effective against one of these is unlikely to prove effective against the other. In this sense, they point out that the common debiasing technique that asks people to reflect on positive exemplars and then make evaluations with those exemplars in mind is indeed likely to mitigate the implicit bias that associates racial categories with negative affect. But, as they emphasise, this same technique is unlikely to prove effective against biases stemming from an association of racial categories with physical constructs (e.g., being rhythmic or athletic). That is because considering positive exemplars of some racial categories will only tend to *reinforce* such associations – e.g., calling to mind the positive examples of Michael Jordan or Beyoncé can entrench association between black people and

---

[18]Many such studies rely on the Implicit Association Test (IAT) – a reaction-time measure that asks people to quickly sort images and words into groups, making as few mistakes as possible. In recent years, the IAT has been heavily criticised. For instance, one of the main concerns raised about the IAT is that it poorly predicts behaviour (Oswald *et al.* 2013; cf. Forscher *et al.* 2019). However, recent theoretical developments seek to put such findings into perspective to show that they are not as problematic as some would have us believe (see, e.g., Brownstein *et al.* 2020).

[19]For a review of some such findings, see Brownstein (2019) and Byrd (2021).

constructs of athleticism and being 'rhythmic', respectively. More broadly, these considerations indicate that adopting debiasing techniques requires a detailed understanding of the mechanisms underwriting implicit biases.

We are now in a position to explain the relevance of these considerations for our purposes. Recall that as discussed at the outset of this section, there is some reason to expect that deliberation can mitigate the deleterious effects of implicit biases. However, the considerations above suggest that we cannot talk of implicit biases as a unified category. And so, there is little reason to think any single debiasing technique will prove effective against *every* kind of implicit bias. In other words, if implicit biases are underwritten by a variety of distinct psychological processes, then any single debiasing technique will be at best partial: even if it proves effective against some kinds of implicit biases, it will not prove effective against a class of others. We should then expect that even if deliberation can help mitigate the effects of some biases, it will nevertheless systematically fail to offset *at least* some *others*. To further buttress this proposal, I will now argue that deliberation fails to correct for a *particular* category of implicit biases.

Most early research on implicit biases relies on the idea that implicitly biased behaviour is predicated on associations. According to this view, implicit biases are underwritten by an implicit association between certain social groups and a given set of features. However, this traditional conception of implicit bias has come under heavy scrutiny in recent years. For example, Del Pinal and colleagues (Del Pinal *et al.* 2017; Del Pinal and Spaulding 2018) argue that some implicit biases are *not* predicated on associations, but are instead encoded in more rich and deeply embedded networks in our concepts. In particular, they contend that some biased behaviour stems from a tendency to regard particular features as *central* to such concepts. Roughly put, a feature f is central to a concept C if many other features of C *depend* on f. To illustrate, consider the concept BIRD. Many of the features we typically associate with BIRD – e.g., flight, laying eggs, and singing – *depend* on them *having a heart*. This is demonstrated by the fact that most people find it hard to disassociate the feature of *having a heart* from their conceptual representations of birds – as they understand that absent a heart, birds would not manifest the other features we typically associate with them (e.g., flight and singing). Thus, *having a heart* is typically regarded as a central feature of our concept BIRD.

In a series of studies, Del Pinal and colleagues (Del Pinal *et al.* 2017) argue that this notion of centrality of features in our concepts helps explain the 'brilliance-gender' bias: that is, the tendency to judge that women are less intellectually brilliant than men. In particular, their studies focus on our concepts MALE PROFESSOR and FEMALE PROFESSOR. Their findings show, perhaps surprisingly, that people associate *smartness* with both these concepts to *a same degree*. However, their findings also show that although people tend to think that features typically associated to male professors – e.g., being knowledgeable, competent, successful – are explained by their smartness, they do not think the same for female professors. Instead, such features of female professors are often taken to be explained by some *other* factor – e.g., their dedication or hard work. Yet another finding from their studies is that people have a hard time disassociating smartness from examples of male professors – similar to how it can be hard to disassociate *having a heart* from examples of birds. However, people found it relatively easier to conceive of female professors who are not smart. Taken together, these findings suggest that although people tend to regard *smartness* as a central feature of MALE PROFESSOR, they do not regard it as central to FEMALE PROFESSOR. Moreover, given that *smartness* is typically associated with both MALE PROFESSOR and FEMALE PROFESSOR, then the gender-brilliance

bias is best explained as arising from a difference in the centrality of *smartness* in these concepts (Del Pinal *et al.* 2017; Del Pinal and Spaulding 2018).

In what follows, I argue that political deliberation cannot correct for biases encoded in central features in our concepts. To begin, recall the three ways in which deliberation can be expected to help alleviate biases (discussed at the outset of this section): (i) deliberation can lead people to become convinced by arguments that are favourable to certain social groups; (ii) people may be prompted to consider scenarios and examples that are counter-stereotypical; and (iii) deliberation can help people take on different perspectives and focus on other (non-stereotypical) features of social groups. I contend that neither of these interventions enabled by deliberation will help correct for biases encoded in central features of conceptual representations. For example, political deliberation does not seem poised to correct the implicit bias against female professors described above. For, as explained above, this bias is not predicated on a lack of an association between female professors and *smartness*. Indeed, as Del Pinal *et al.* (2017) demonstrate, people strongly associate *smartness* with female professors. As such, it would be ineffective to present people afflicted with this bias with persuasive arguments to the effect that female professors are smart (or even smarter than men), nor would it help to prompt them to focus on this individual feature of female professors. Since *smartness* is *already* regarded as a typical feature of female professors, such strategies would not target the underlying mechanisms giving rise to this bias. Rather, what seems to be required is a lengthy and dedicated intervention geared towards changing the many deep-rooted dependency networks of the concept FEMALE PROFESSOR. And presumably, ordinary political deliberation does not *itself* provide conditions for this to occur.

To further buttress these arguments, consider once more the scenario described in *Quota* (described earlier in this section). Suppose that the implicit bias afflicting members of the deliberating collective C is due to their implicit tendency to regard *aggressiveness* as a central feature of the concept IMMIGRANT. Presumably, deliberation would also not help to mitigate this bias. For, if aggressiveness is a central feature of IMMIGRANT, then it would be hard to disassociate it from this concept – similar to how it would prove difficult to disassociate the feature of having hearts from our concept BIRD. As such, it would not help to merely present members of C with persuasive arguments favourable to immigrants since any example of immigrant they consider would have the feature of aggressiveness. Neither would it help to prompt them to consider examples of immigrants who were purportedly not aggressive. If prompted to do so, they would most likely regard such examples as merely fictitious – similar to how the notion of a heartless bird strikes many as merely fictitious. Moreover, asking members of C to focus on other individual features of immigrants would also prove ineffective: for, if aggressiveness is a central feature of IMMIGRANT, then it will most likely be connected to many other features of immigrants. And so, considering these features will fail to correct for the deep-rooted dependency network underwriting this bias.

More broadly, I take these considerations to show that ordinary political deliberation most likely systematically fails to correct for biases encoded in central features of our concepts. After all, none of the reasons given for thinking deliberation has debiasing potential seem to give us reason it will correct for this class of implicit biases. Mitigating such biases seem to require instead lengthy and dedicated interventions geared towards changing the many deep-rooted dependency networks of such concepts. These arguments leave it open that deliberation conducted by experts might prove more effective in doing so. But what is crucial to point out is that typical political deliberation *itself* does not seem to provide the conditions for this to occur.

3.3. Taking stock

In this section I first argued the most prominent epistemic arguments for deliberative democracy are conditional on whether deliberation is more likely to mitigate the deleterious effects of implicit biases than alternative methods of political decision-making. I then explained that deliberation will most likely systematically fail to correct for at least some classes of implicit biases – e.g., biases encoded in central features of our concepts. These claims suggest that all the most prominent epistemic arguments for deliberative democracy overstate the epistemic benefits of political deliberation. After all, instrumental epistemic arguments claim that deliberative democracy is more likely to produce correct political decisions when compared with alternative models of political decision-making. But, plausibly, for cases where implicit biases encoded in central features of concepts permeate deliberation, deliberative democracy will prove no more likely to promote correct political decision-making than majority voting. Moreover, if these kinds of implicit biases permeate deliberation, then deliberative democracy is no more likely than majority voting to foster the formation of political beliefs in response to the evidence, or to promote mutual epistemic accountability – thus challenging non-instrumental epistemic arguments. Taken together, these considerations suggest that the most prominent formulations of epistemic arguments for deliberative democracy are unwarranted.

There are two ways proponents of epistemic arguments for deliberative democracy may seek to respond to the above claims. The first is to contend that the arguments here stand in need of critical empirical support. Indeed, I have only given reasons to think deliberation *might* systematically fail to correct for some implicit biases (specifically, biases encoded in central features of our concepts). However, it is true that there has not yet been any dedicated empirical investigation of this proposal. And so, the arguments developed in this section are not decisive.

In reply, I maintain that proponents of epistemic arguments for deliberative democracy are not in the position to simply pass on the burden of proof to the opposition in this way. After all, their claims also stand in need of critical empirical support. As previously explained, these arguments are conditional on the potential of deliberation to mitigate implicit biases. But there is a dearth of empirical evidence showing that deliberation proves effective in this regard. And insofar as the findings surveyed in this section provide reason to be suspicious of this hypothesis, then there seems to be stronger reasons to favour the sceptical position defended here.

A second line of reply they can advance is to claim that, *on average*, political deliberation is still more epistemically virtuous than relevant alternative methods of political decision-making. That is, they might argue that even if in *some cases* political deliberation is no more epistemically virtuous than, say, majority voting, it is still epistemically better in a greater number of other cases. And so, on balance deliberative democracy is justified. This response is seemingly plausible. However, it relies on one of two claims: either that implicit biases do not permeate most cases of political decision-making; or that for most cases of political decision, when implicit biases permeate the decision-making process, those biases are of a sort that deliberation proves effective at mitigating. But there is no empirical evidence in support of either of these claims. As such, I take this response to be ultimately unmotivated.

4. A New Way Forward

In this last section, I explain how the arguments in this paper prove instructive for future work on the normative ideal of deliberative democracy. My proposal is that

the kinds of empirical findings and methods I deployed in those arguments can aid in the development of a new (and improved) generation of epistemic arguments for deliberative democracy.

To begin, it is useful to briefly outline the two major positions in recent debates surrounding epistemic arguments for deliberative democracy. On one side of these debates are the optimists. They claim that findings about the cognitive benefits of deliberation warrant the sweeping conclusion that *political* deliberation will improve political decision-making across the board (Mercier and Landemore 2012; Landemore 2017; Mercier and Sperber 2017). On these grounds, optimists argue that we have enough reason to favour deliberative democracy. For instance, Landemore (2017) relies on developments in evolutionary psychology to argue that under 'normal' circumstances, deliberation *will* outperform individual reasoning on political issues – and, therefore, that deliberation is an essential component of the democratic process. On the other side of this debate are the pessimists. They argue that evidence showing how deliberation fails to improve decision-making in certain cases warrants the sweeping conclusion that we should be sceptical about deliberative democracy (Hibbing and Theiss-Morse 2002; Sunstein 2007).

To a first approximation, the arguments from the previous sections might be taken to favour the pessimist view. After all, if deliberation systematically fails to correct for implicit biases, then this seems to favour scepticism about epistemic defences of deliberative democracy. However, I will now argue that both the optimist and the pessimist positions are problematic. The sweeping generalisations optimists and pessimists seek to advance from small samples of empirical evidence to either wholesale rejection or acceptance of deliberative democracy are unjustified.

Recall findings discussed in the previous section suggesting that different implicit biases stem from distinct psychological processes. Whereas some implicit biases can be encoded in simple associations, others are encoded in more rich and deeply embedded networks in our concepts. This underscores that there is no single category of implicit bias – and that phenomena subsumed under this label form a heterogeneous bunch. Furthermore, as we also saw, this heterogeneity suggests that techniques poised to correct for one kind of implicit bias might prove entirely ineffective against others. I contend that these findings render invalid any attempt to generalise from extant evidence about particular (de)merits of deliberation to comprehensive assessments of deliberative democracy. After all, even if empirical findings demonstrate that deliberation can prove effective at mitigating certain kinds of biases, this offers no guarantee that it will succeed in improving political decision-making across the board. Indeed, if the arguments in the previous section are correct, then it is questionable that deliberation will correct for biases encoded in central features of our concepts. In this respect, the optimist position fails. Yet, this does not suffice to show that deliberation should not be an essential component of the political decision-making process under certain circumstances. After all, deliberation may have significant epistemic benefits in at least some circumstances: even if deliberation does not correct for biases encoded in central features of our concepts, there is still reason to think they can correct for other kinds of biases. This then provides reason to reject the pessimist position as well. In sum, attempts to simply defend or reject deliberative democracy via a generalisation from the small bodies of empirical evidence on whether deliberation improves/ does not improve decision-making are unwarranted. The tendency of either camp to favour these sweeping generalisations is at odds with empirical reality.

In contrast to these extreme camps, I contend that future work on the normative ideal of deliberative democracy should attend to the *particular* circumstances under

which deliberation is indeed poised to improve political decision-making. In this respect, I contend that research on epistemic defences of deliberative democracy should turn to a piecemeal investigation of which *specific* implicit biases deliberation can correct and which ones it cannot. For this task, normative theorists should engage with empirical findings and methods from cognitive and social psychology (in line with what I developed in the previous section). The hope is that this would allow for a better-informed evaluation of when we should adopt deliberative practices in political decision-making and when we should refrain from adopting such methods. In other words, by attending to the psychological processes underwriting implicit biases, normative theorists would be able to develop more precise and empirically informed epistemic arguments for deliberative democracy that can better guide political decision-making. I lack the space to develop these suggestions in great detail here; but I take the arguments in the previous section as sufficiently instructive to initiate the development of a new generation of normative epistemic arguments for deliberative democracy that can move us beyond the stale dialectic of the extreme pessimist and optimist positions.

5. Conclusion

I began by outlining the main tenets of the most prominent formulations of epistemic defences of deliberative democracy. I then developed a two-step argument for thinking these epistemic defences are unwarranted. First, I showed that these arguments are conditional on the potential of political deliberation to mitigate the deleterious effects of implicit biases. And second, I argued that recent work on implicit biases suggests that political deliberation will systematically fail to correct for some classes of implicit bias. I then explained why this does not thereby give us reason to be outright sceptical about the prospects of deliberative democracy. My main contention was that findings and methods from cognitive and social psychology can aid in the development of a new and improved generation of epistemic arguments for deliberative democracy – one that attends to the particular circumstances under which deliberation can indeed improve political decision-making, rather than relying on unwarranted sweeping generalisations about the epistemic virtues of political deliberation.[20]

References

Alston W.P. (1989). *Epistemic Justification: Essays in the Theory of Knowledge*. Ithaca, NY: Cornell University Press.

Anderson E. (2006). 'The Epistemology of Democracy.' *Episteme* **3**(1–2), 8–22.

Bächtiger A., Dryzek J.S., Mansbridge J. and Warren M.E. (eds) (2018). *The Oxford Handbook of Deliberative Democracy*. Oxford: Oxford University Press.

Benhabib S. (1996). 'Toward a Deliberative Model of Democratic Legitimacy'. In S. Benhabib (ed.), *Democracy and Difference: Contesting the Boundaries of the Political*, pp. 67–94. Princeton: Princeton University Press.

Blair I.V., Ma J.E. and Lenton A.P. (2001). 'Imagining Stereotypes Away: The Moderation of Implicit Stereotypes Through Mental Imagery.' *Journal of Personality and Social Psychology* **81**(5), 828–41.

Bohman J. (1998). 'Survey Article: The Coming of Age of Deliberative Democracy.' *Journal of Political Philosophy* **6**(4), 400–25.

Bohman J. (2007). 'Political Communication and the Epistemic Value of Diversity: Deliberation and Legitimation in Media Societies.' *Communication Theory* **17**(4), 348–55.

Brinol P., Petty R. and McCaslin M. (2009). 'Changing Attitudes on Implicit Versus Explicit Measures: What Is the Difference?' In R. Petty, R. Fazio and P. Brinol (eds), *Attitudes: Insights from the New Implicit Measures*, pp. 305–46. New York, NY: Psychology Press.

Brownstein M. (2019). 'Implicit Bias.' In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Fall 2019 edition. https://plato.stanford.edu/archives/fall2019/entries/implicit-bias/.

Brownstein M., Madva A. and Gawronski B. (2020). 'Understanding Implicit Bias: Putting the Criticism into Perspective.' *Pacific Philosophical Quarterly* **101**(2), 276–307.

Brownstein M. and Saul J. (2016). *Implicit Bias and Philosophy, Volume 2: Moral Responsibility, Structural Injustice, and Ethics*. Oxford: Oxford University Press.

Buckwalter W. (2019). 'Implicit Attitudes and the Ability Argument.' *Philosophical Studies* **176**(11), 2961–90.

Byrd N. (2021). 'What We Can (and Can't) Infer about Implicit Bias from Debiasing Experiments.' *Synthese* **198**(2), 1427–55.

Cohen J. (2003). 'Deliberation and Democratic Legitimacy'. In D. Matravers and J. Pike (eds), *Debates in Contemporary Political Philosophy*, pp. 67–92. London: Routledge.

Cohen J. (2009). 'Truth and Public Reason.' *Philosophy & Public Affairs* **37**(1), 2–42.

Del Pinal G., Madva A. and Reuter K. (2017). 'Stereotypes, Conceptual Centrality and Gender Bias: An Empirical Investigation.' *Ratio* **30**(4), 384–410.

Del Pinal G. and Spaulding S. (2018). 'Conceptual Centrality and Implicit Bias.' *Mind & Language* **33**(1), 95–111.

Estlund D.M. (2008). *Democratic Authority: A Philosophical Framework*. Princeton, NJ: Princeton University Press.

Estlund D.M. and Landemore H. (2018). 'The Epistemic Value of Democratic Deliberation.' In A. Bächtiger, J.S. Dryzek, J. Mansbridge and M.E. Warren (eds), *The Oxford Handbook of Deliberative Democracy*, pp. 113–31. Oxford: Oxford University Press.

Forscher, P.S., Lai C.K., Axt J.R., Ebersole C.R., Herman M., Devine P.G. and Nosek B.A. (2019). 'A Meta-Analysis of Procedures to Change Implicit Measures.' *Journal of Personality and Social Psychology* **117**(3), 522–59.

Galinsky A.D. and Moskowitz G.B. (2000). 'Perspective-Taking: Decreasing Stereotype Expression, Stereotype Accessibility, and in-Group Favoritism.' *Journal of Personality and Social Psychology* **78**(4), 708–24.

Gaus G. (1996). *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. Oxford: Oxford University Press.

Goodin R.E. and Spiekermann K. (2018). *An Epistemic Theory of Democracy*. Oxford: Oxford University Press.

Gutmann A. and Thompson D.F. (2009). *Why Deliberative Democracy?* Princeton, NJ: Princeton University Press.

Habermas J. (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: Polity Press.

Hermanson S. (2017). 'Rethinking Implicit Bias: I Want My Money Back.' https://leiterreports.typepad.com/blog/2018/04/sean-hermanson-rethinking-implicit-bias-i-want-my-money-back.html.

Hibbing J.R. and Theiss-Morse E. (2002). *Stealth Democracy: Americans' Beliefs about How Government Should Work*. Cambridge: Cambridge University Press.

Holroyd J. and Sweetman J. (2016). 'The Heterogeneity of Implicit Bias.' In M. Brownstein and J. Saul (eds), *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*, pp. 80–103. Oxford: Oxford University Press.

Horwich P. (1998). *Truth*. Oxford: Clarendon Press.

Ichikawa J.J. and Steup M. (2018). 'The Analysis of Knowledge.' In E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, Summer 2018 edition. https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/.

Landemore H. (2017). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton, NJ: Princeton University Press.

Machery E. (2016). 'De-Freuding Implicit Attitudes.' In M. Brownstein and J. Saul (eds), *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*, pp. 104–29. Oxford: Oxford University Press.

Machery E. (2017). 'Should We Throw the IAT on the Scrap Heap of Indirect Measures?' https://philoso-phyofbrains.com/2017/01/17/how-can-we-measure-implicit-bias-a-brains-blog-roundtable.aspx.

Madva A. (2017). 'Biased Against Debiasing: On the Role of (Institutionally Sponsored) Self-Transformation in the Struggle Against Prejudice.' *Ergo: An Open Access Journal of Philosophy* **4**. doi: 10.3998/ergo.12405314.0004.006.

Madva A. and Brownstein M. (2018). 'Stereotypes, Prejudice, and the Taxonomy of the Implicit Social Mind.' *Noûs* **52**(3), 611–44.

Mansbridge J., Bohman J., Chambers S., Estlund D., Føllesdal A., Fung A., Lafont C., Manin B. and Martí J.L. (2010). 'The Place of Self-Interest and the Role of Power in Deliberative Democracy.' *Journal of Political Philosophy* **18**(1), 64–100.

Martí J.L. (2006). 'The Epistemic Conception of Deliberative Democracy Defended.' In S. Besson and J.L. Martí (eds), *Democracy and Its Discontents: National and Post-National Challenges*, pp. 27–56. Hampshire: Ashgate.

Mercier H. and Landemore H. (2012). 'Reasoning is for Arguing: Understanding the Successes and Failures of Deliberation.' *Political Psychology* **33**(2), 243–58.

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

Misak C. (2002). *Truth, Politics, Morality: Pragmatism and Deliberation*. London: Routledge.

Nino C.S. (1996). *The Constitution of Deliberative Democracy*. New Haven, CT: Yale University Press.

Nosek B.A., Greenwald A.G. and Banaji M.R (2007). 'The Implicit Association Test at Age 7: A Methodological and Conceptual Review'. In J. A. Bargh(ed), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes*, pp. 265–92. New York: Psychology Press.

Oswald F., Mitchell G., Blanton H., Jaccard J. and Tetlock P. (2013). 'Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies.' *Journal of Personality and Social Psychology* **105**. https://doi.org/10.1037/a0032734.

Peter F. (2009). *Democratic Legitimacy*. London: Routledge.

Peter F. (2013). 'The Procedural Epistemic Value of Deliberation.' *Synthese* **190**(7), 1253–66.

Peter F. (2016). 'The Epistemic Circumstances of Democracy'. In M. Brady (ed), *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*, pp. 133–49. New York: Oxford University Press.

Pettigrew R. (2016). *Accuracy and the Laws of Credence*. Oxford: Oxford University Press.

Rawls J. (2005). *Political Liberalism*. Ithaca, NY: Columbia University Press.

Schwartzberg M. (2015). 'Epistemic Democracy and Its Challenges.' *Annual Review of Political Science* **18**, 187–203.

Sunstein C.R. (2007). 'Group Polarization and 12 Angry Men.' *Negotiation Journal* **23**(4), 443–7.

Talisse R.B. (2009). *Democracy and Moral Conflict*. Cambridge: Cambridge University Press.

Urbinati N. (2014). *Democracy Disfigured*. Cambridge, MA: Harvard University Press.

Young I.M. (1996). 'Communication and the Other: Beyond Deliberative Democracy.' *Democracy and Difference: Contesting the Boundaries of the Political* **31**, 120–35.Notes

Miguel Egler is an Assistant Professor in Philosophy at Tilburg University, where he is a member of the Tilburg Center for Moral Philosophy, Epistemology and Philosophy of Science.