ON THE EPISTEMIC VALUE OF REFLECTION

Forthcoming in *Ergo* (This is a penultimate draft; please cite the published version)

PRANAV AMBARDEKAR
The Ohio State University

Against philosophical orthodoxy, Hilary Kornblith has mounted an empirically grounded critique of the epistemic value of reflection. In this paper, I argue that this recent critique of the epistemic value of reflection fails even if we concede that (a) the empirical facts are as Kornblith says they are and (b) reliability is the only determinant of epistemic value. The critique fails because it seeks to undermine the reliability of reflection *in general* but targets only one of its variants, namely *individual reflection*, while neglecting *social reflection*. This critique comprises two arguments which have a common structure: they both impose a requirement on the reliability of reflection, but deny, on empirical grounds, that the requirement is met. One argument imposes an introspection requirement, which I reject as superfluous. I show how reflection can proceed without introspection. The other argument imposes an efficacy requirement. This requirement concerns whether reflection is causally efficacious i.e., whether it leads us to change our minds for the better. I accept this as a genuine requirement. Even if we concede that *individual reflection* fails to meet this requirement, I argue that we have not been given sufficient evidence to believe that *social reflection* is bound to fail this requirement. Furthermore, my analysis of the conditions under which *social reflection* works best provides us with *prima facie* grounds for optimism regarding the reliability of *social reflection*. Ultimately, then, these arguments fail to undermine the epistemic value of reflection in general.

Western philosophers throughout the ages have held reflection in high regard. Ancient Greek philosophers like Socrates and Aristotle thought that reflection was central to the good life. Enlightenment and post-Enlightenment philosophers emphasized the epistemic value of reflection. In Descartes' (1990) work we find the idea that beliefs which survive critical reflection are thereby more rational and those that fail in this regard are rationally defective. Mill (2009), Clifford (1999), and Nietzsche (1974) added that we are obliged, qua *believers*, to subject our beliefs to critical scrutiny. For these philosophers, critically scrutinizing one's beliefs involves reassessing one's evidence, developing objections against our beliefs, and so on. Mill (2009: 63) remarks that such critical reflection is central to "real understanding." These ideas, and the general outlook behind them, have exerted a considerable influence on contemporary epistemology.[1]

This sanguine outlook on the epistemic value of reflection has recently been challenged. Kornblith (2012) has mounted an empirically grounded critique of the epistemic value of reflection. Kornblith takes reliability as the only determinant of epistemic value. In other words, he thinks that some mechanism or process is epistemically valuable only if it reliably leads to the truth (2012: 34). Using results from empirical psychology, he argues against the reliability of reflection. Without reliability, reflection loses its epistemic value. And so Kornblith maintains that "philosophers have typically assigned a great deal more value to reflection than it deserves." (2012: 1) The critique purports to upset a long-standing tradition of philosophical thinking.

In this paper, I argue that this critique fails even if we grant that (a) the empirical facts are as Kornblith says they are and (b) reliability is the only determinant of epistemic value. The

[1] See Joshi (2021) on Socrates, Aristotle, Mill, and Nietzsche's views on the centrality of reflection to the good life. See also Meritt (2018) on the importance of reflection in Kant's epistemology. For discussion on the connection between reflection and rational belief in contemporary epistemology, see Bonjour (1985), Alston (1989), Burge (1996), Kornblith (2012), and Smithies (2015).

critique fails because it seeks to undermine the reliability of reflection *in general* but targets only one of its variants, namely *individual reflection*, while neglecting *social reflection*. Philosophers who appeal to reflection have often been accused of having an overly individualistic epistemology, à la Descartes. In recent years, epistemology has taken a social turn. Critics of Cartesian approaches to epistemology have emphasized the epistemic significance of testimony and disagreement. But reflection has a social dimension as well. Some critics of reflection seem to have underestimated this point.

Reflection is defined here as *second-order reasoning*: reasoning that is guided by second-order considerations about one's beliefs or one's normative reasons for belief.[2] So defined, reflection includes two variants: individual reflection and social reflection. An agent reflects *individually* when they are deliberating on normative reasons for belief without collaborating with other agents. It is a mistake to think that this is all there is to reflection. For reflection is often done in a *social* context as well, as when we play what Brandom (1994) calls "the game of giving and asking for reasons." Sometimes, when an agent asserts her belief that p, they are confronted by an interlocutor who disagrees with them. As the agents pursue their disagreement, they deliberate about what good reasons there are for believing that p. In doing so, they have begun reflecting *socially*.

This distinction is crucial for responding to the critique. The critique of reflection comprises two independent arguments: *The Introspection Argument* and *The Efficacy Argument*. These arguments have a common structure. They both impose a requirement on the reliability of reflection, but deny, on empirical grounds, that the requirement is met. I shall argue that *The*

[2] This definition is not meant to capture all the uses of the word "reflection" in ordinary language or theoretical discourse. In philosophical literature, the theoretical term "deliberation" has often been used to refer to what I call reflection. See for instance, Shah and Velleman's (2005) gloss on "doxastic deliberation." However, Shah and Velleman's discussion is limited to individual reflection.

Introspection Argument imposes a dubious requirement which can be rejected on independent grounds – that is, without reference to social reflection. However, *The Efficacy Argument* imposes a genuine requirement on the reliability of reflection. Although Kornblith argues that reflection violates this requirement, his argument does not extend beyond individual reflection. Ultimately, social reflection escapes his critique of the epistemic value of reflection.

To be clear, this paper does not provide a *positive* argument for the overall reliability of reflection. Instead, it defends the epistemic value of reflection against Kornblith's *negative* argument against the overall reliability of reflection. Whether or not his critique impugns individual reflection – and here I remain officially neutral – I show that it fails to impugn social reflection. As a result, it fails to undermine the overall reliability of reflection.

Here is how the paper is organized. In section 1, I clarify the notion of reflection, and present some contemporary views of the epistemic value reflection. In section 2, I reject *The Introspection Argument* on the grounds that it imposes a superfluous requirement on the reliability of reflection. In section 3, I reject *The Efficacy Argument* on the grounds that we lack sufficient evidence that social reflection violates its requirement on the reliability of reflection. In section 4, I conclude that Kornblith's critique fails, whilst acknowledging that future work may need to address other empirically grounded critiques of the epistemic value of reflection.

1. The Epistemic Value of Reflection

1.1. What is Reflection?

To a first approximation, reflection is a capacity to double-check our unreflective responses to the world, including beliefs and intentions. Reflection – as I understand it – refers to a process of second-order reasoning wherein agents seek to confirm or revise their unreflective responses in light of normative considerations about the reasons for or against those responses.

For example, when I make the perceptual judgment that there is a cat in front of me, I might ask myself, "Is my perceptual experience a good reason to believe that there is a cat in front of me? I am, after all, in the museum of illusions." Consequently, I might arrive at the second-order belief that my first-order judgment about the existence of the cat is unjustified. Reflection is thus a higher-order mental phenomenon. Following Pettit (2007: 498-500), there are two distinct kinds of higher-order reasoning: one could reason *meta-attitudinally* or *meta-propositionally*.[3] The former involves reasoning about one's own first-order attitudes, as in the museum of illusions example. The latter involves reasoning about evidential support relations, as when I ask myself the question, 'Is p a good reason to believe q?' Such reasoning concerns relations between the propositional contents of mental states rather than the mental states themselves.

Like me, Kornblith (2012: 1, 28, 43, 109) understands reflection as a process of "second-order scrutiny" guided by normative considerations, and not just any kind of careful thinking. For instance, he writes:

[3] Peacocke (1996) makes a similar distinction between *self-directed* and *world-directed* modes of critical reasoning in his reply to Burge (1996).

> We are capable of reflecting on our beliefs and desires, stopping to assess them, and stopping to question the wisdom and advisability of further belief or action. When we reflect, and when we engage in reflective evaluation of our first-order states, it seems that we bring to bear certain normative standards on our beliefs and actions in a way that other animals could not possibly do. (2012: 109)

As evidenced by his discussion of various cases, Kornblith (2012: 20, 89, 142) acknowledges not only meta-attitudinal reasoning, but also meta-propositional reasoning as a genuine form of reflection. Consider one such case:

> *Jury.* Suppose that I am serving on a jury in which someone is charged with murder. Imagine as well that I don't simply react to the evidence presented. Instead, I stop to reflect. I self-consciously consider whether the evidence presented supports a guilty verdict. (2012: 89)

In *Jury*, the agent does not reflect on their own first-order beliefs about the matter. Instead, the agent reflects on whether the evidence presented to them supports a particular proposition, namely that the accused is guilty. The agent's reflection is meta-propositional, not meta-attitudinal.

Now, it is crucial to note that on this definition of reflection, it is *inessential* whether an agent reflects by herself or does so in collaboration with others. For example, thinking through a logic quiz can be an instance of reflection, whether one does so alone or jointly with other quizzers. And one's reflection could be meta-attitudinal (reasoning about our own beliefs about the quiz) or

meta-propositional (reasoning about the validity of steps in a proof).[4] Social reflection is *collaborative* in nature, requiring the joint effort of multiple agents. Consequently, the quizzers' reflection is an instance of social reflection only if they work through the logic quiz *collaboratively*.

Philosophers influenced by Descartes have tended to hold a narrow conception of reflection on which reflection just means meta-attitudinal reasoning restricted to the individual case. This conception is captured in the familiar picture of the lone Cartesian thinker who takes a step back from their inner life and asks whether the first-order attitudes they have are worth having (Frankfurt 1971; Korsgaard 1996). Even if it is the case that Kornblith's arguments undermine only this narrow conception of reflection, it would be a mistake to think that the narrow conception was his intended target all along. For Kornblith's (2012: 89, 109) conception of reflection is broader than the Cartesian one.

As it turns out, then, Kornblith fails to recognize the implications of the broad conception of reflection he endorses. More specifically, he overlooks the fact that meta-attitudinal and meta-propositional reasoning can be done socially as well as individually. As we shall see, this distinction is crucial in rebutting his critique of the epistemic value of reflection.

1.2. Why is Reflection Epistemically Valuable?

Philosophers diverge on what reflection can achieve, and the epistemic value it generates. Philosophers from the Harvard-Pittsburgh tradition cash out the epistemic value of reflection in

[4] Often, the entire process of working on a logic quiz – whether alone or in a group – incorporates a mix of both meta-attitudinal and meta-propositional reasoning.

terms of the basing relation that is necessary for knowledge or justified belief (Korsgaard 1996; Moran 2001). They hold that in order for an agent to believe that p on the basis that q, (i) the agent must have the higher-order reflective capacity to represent that q is a good reason for p, and (ii) that such a capacity is relevant to explaining why the agent holds the belief that p. *Epistemic basing* is a necessary condition on justified or rationally held belief: if one has a justified belief that p, then one believes p *on the basis of* good reasons. These philosophers think that without reflection we cannot have justified belief. They differ on how exactly reflection manages to establish basing relations. Some, like Moran (2001) and Leite (2004), argue that sincere avowals – such as "I really think that p because q" – can directly establish basing relations. Since knowledge requires justified belief, these philosophers are committed to saying that we cannot have knowledge without reflection either. Both knowledge and justified belief are said to have epistemic value. For many epistemologists, knowledge has epistemic value *par excellence*. Some epistemologists cash out the epistemic value of justified or rationally held belief without reference to knowledge (Wedgwood, 2017). In the Harvard-Pittsburgh tradition, reflection is epistemically valuable to the extent that it generates knowledge or justified belief.

This approach has been accused of over-intellectualizing the basing relation, and along with it, epistemic statuses like rational belief and knowledge (Kornblith 2012: Ch. 2). Opponents argue that some pre-reflective and non-reflective agents (human infants and animals respectively) can have both rational belief and knowledge. When a squirrel forms the belief that 'here is an acorn,' for example, it has responded rationally to its perceptual evidence, which in turn gives it reason for belief. And so the squirrel can come to know that 'here is an acorn.' The squirrel displays *sensitivity to reasons*, which is necessary for rational belief and knowledge. The capacity for reflection is not required for either epistemic status.

To avoid the over-intellectualizing problem, some epistemologists have argued that reflection is a necessary condition on *epistemic responsibility*, as opposed to justified belief or knowledge (Burge 1996; Smithies 2015). The potential to reflect on the justification of one's beliefs opens up the agent's management of one's doxastic life to public criticism. Such an agent is a fair target for *epistemic* (as opposed to moral or pragmatic) praise or blame. In holding someone epistemically responsible for their beliefs, one makes a demand on them to comply with certain normative standards. This practice has two presuppositions. First, that the target can understand our demand. Second, that the target can bring this understanding to bear on causally regulating their own beliefs.

There are various accounts of the epistemic value of *holding beliefs in an epistemically responsible way*. One strategy is to say that beliefs that are held in an epistemically responsible way tend to be more reliable: epistemic responsibility is truth-conducive (Smithies 2015). The thought is that first-order beliefs that pass the test of reflective scrutiny are more likely to be true.

Whatever one thinks about epistemic responsibility and why it is epistemically valuable, several contemporary epistemologists will agree with the following minimal commitment: even if reflection is not required for epistemic responsibility, it often leads to true belief. In slogan form: reflection is a *tolerably* reliable method of belief formation and maintenance. Indeed, something like this minimal commitment might be required to explain philosophers' enthusiasm for promoting critical thinking courses in university and high school settings.

To sum up, contemporary epistemologists have variously thought that reflection is necessary for (a) rational belief, (b) knowledge, or (c) epistemic responsibility. Those who disagree that reflection is required for (a) and (b), but hold that reflection is required for (c) cash out its value in terms of epistemic reliability, or in some other way. And pretty much every epistemologist

will agree that forming beliefs via reflection, or maintaining them by putting them under reflective scrutiny, sometimes increases their reliability.

2. The Introspection Argument

In the last decade or so, philosophers like Kornblith (2012: Ch. 1 & Ch. 5) and Doris (2015: Part I) have mounted an empirically grounded critique of the epistemic value of reflection. A closer look at their critiques reveals two independent arguments aimed at undermining the reliability of reflection: *The Introspection Argument* and *The Efficacy Argument*. For the purposes of this paper, I shall understand "the critique of the epistemic value of reflection" to consist of these two arguments. In this section, I shall critically evaluate *The Introspection Argument*.

The Introspection Argument

1. **Epistemic Value Reliabilism**: The only determinant of epistemic value is reliability.
2. Therefore, reflection has epistemic value only insofar as it is reliable.
3. **Introspection Requirement**: Reflection is reliable only if we can reliably identify the reasons for which we hold our beliefs.
4. **Anti-Introspection**: Psychology shows that we cannot reliably identify the reasons for which we hold our beliefs.
5. Therefore, reflection is unreliable.
6. Therefore, reflection has little or no epistemic value.

Epistemic Value Reliabilism is controversial. An epistemic value pluralist can reject this premise by cashing out the value of reflection in terms of rationally held belief, intellectual autonomy, personhood, or some other determinant of epistemic value.[5] Both the arguments in the critique of reflection presuppose **Epistemic Value Reliabilism**. For the purposes of this paper, I shall grant this premise. This should not worry epistemic value pluralists. Whatever else reflection may get us, I take it that everyone is interested in preserving the idea that reflection is a tolerably reliable mechanism of belief formation and maintenance.

My discussion will focus on **Anti-Introspection** and the **Introspection Requirement** – the two key premises of this argument. **Anti-Introspection** is an empirical thesis which states that the **Introspection Requirement** is seldom met. Why should anyone accept **Anti-Introspection**? In what follows, I review a range of empirical evidence cited by Kornblith (2012: 20-26) and Doris (2015: 41-100) in support of **Anti-Introspection**.

2.1. The Empirical Evidence for Anti-Introspection

Studies from empirical psychology provide evidence for the following phenomena:

> **Opacity**: We often form and maintain our beliefs due to influences that we are unaware of; influences that are opaque to first-person introspection.

[5] For instance, see Smithies (2019: 279-282) on the connection between reflection, personhood, and epistemic responsibility.

> **Confabulation**: Verbal expressions of one's reasons for belief, however sincere, are often confabulatory: the agent "generates" a normative story to justify their beliefs, such that this story has little to do with why they actually hold the relevant beliefs.

Opacity and **Confabulation** jointly constitute evidence for **Anti-Introspection**. We begin with **Opacity**. Although some of the empirical research I shall discuss below is explicitly only about subliminal influences on action, we can extend the moral of these empirical studies to judgment and belief, since patterns of behavior are often associated with particular judgments and beliefs.

A study conducted by Nisbett and Wilson (1977) revealed that subjects' judgments about the quality of stockings depended significantly on the position of the stockings in the store: in an array of stockings, the stockings placed on the right end of the array were chosen four times more than the stockings placed on the left end. When asked about their reasons for their choices, subjects could only cite features like knit, weave, workmanship etc. The *actual basis* of their judgment, namely the relative position of the stockings, was inaccessible to introspection. Some other prominent examples of arbitrary, subliminal influences on belief and behavior involve pictures of watching eyes (Haley and Fessler 2005), pronouns (Gardner et al. 1999) and the order of names in ballots (Webber et al. 2014).

Kahneman and Tversky's (1982) discussion of the "Asian Disease Problem" suggests that moral judgment can be affected by "ethically arbitrary factors" like how a hypothetical moral problem is verbally framed. In a hypothetical epidemic scenario, participants elicited different responses to a proposed intervention. When the expected outcome of the intervention is framed in terms of survival, people have risk-averse judgments; when the same outcome is framed in terms of expected mortality, people have risk-seeking judgments. Eskine et al. (2011) and Kelly (2011)

have shown that emotions like disgust can affect moral judgment. The framing of a moral problem and certain emotional reactions are obviously *wrong kinds of reasons* for moral judgment/belief. In addition, people rarely, if ever, are disposed to cite such factors as reasons for belief.

Last, there is a vast literature on implicit bias or implicit cognition.[6] Implicit bias involves the pre-reflective attribution of negative qualities to members of social out groups, by race, age, gender, sexuality, weight, and religion. Even if such bias is best modelled using conceptual categories other than belief, the effects of implicit bias appear at the level of both action and belief. Consider implicit racial bias. A study conducted by Bertrand and Mullianathan (2003) showed that fabricated résumés sent in response to employment advertisements in Boston and Chicago newspapers elicited fifty percent more interview callbacks if they contained "white sounding" names as compared to "African American sounding" names. This is clearly a case of racial bias affecting action, but there may be associated beliefs here too i.e., beliefs accompanying racist actions, like "Jamal does not seem to be a guy who is fit for this job."

Now, it is not as if subjects in these studies will stay quiet or remain baffled about their reasons for action or belief. On the contrary, subjects typically produce very confident answers when prompted. **Confabulation** documents a specific instance of a more general phenomenon which goes by the same name: confabulation, understood generally, concerns sincere, confident but erroneous reporting about various aspects of oneself, including one's reasons for action or belief. According to Doris (2015: 82), prominent researchers like Nisbett and Wilson (1977: 233), Gazzaniga (2000: 1316-1321), Hirstein (2005: Ch. 1), and Carruthers (2009: 126-127) all assert

[6] For a comprehensive and accessible discussion on implicit bias, see Banaji and Greenwald (2013).

that confabulation, understood generally, is not restricted to patients with pathological disorders, but is common among healthy people.

A study done by Estabrooks (1957: 86-87) showed that subjects always found socially acceptable ways to justify or excuse their behavior when "incomprehensible behavior" was induced into them via hypnosis. Participants in a study conducted by Maier (1931) were asked to solve a problem. Those who had difficulty finding a solution were given a "hint." Even when the hint was helpful, participants did not credit it for their success. Instead, they said the solution "dawned" on them, or credited their past courses of study. A study done by Latané and Darley (1970) on the so-called "Group Effect" demonstrates the tenacity with which people hold onto their confabulations. "Group Effect" refers to the inversely proportional relationship between a tendency in individuals to extend help and the number of bystanders present. Despite being challenged, participants in the study denied that bystanders had anything to do with their behavior.

For the sake of argument, I shall grant my opponents' assessment of the empirical evidence in favor of **Opacity** and **Confabulation**. Therefore, by extension, I shall grant **Anti-Introspection**. I recognize that some philosophers and scientists may push back here. Philosophers have an important role to play in clarifying the concepts and methods used in empirical research. Sometimes, the upshot is that some strand of empirical research is orthogonal to a long-standing philosophical discussion or debate. When some empirical research is relevant to a philosophical debate, philosophers can point out hasty generalizations based on that research. Scientists might have their own concerns: measurement problems, replicability, data dredging, publication bias, and more specific concerns regarding the methods and assumptions of individual studies. I shall keep all such skepticism at bay. To block *The Introspection Argument*, my strategy is instead to reject the **Introspection Requirement.**

2.2. Rejecting the Introspection Requirement

On a popular way of understanding epistemic basing, the reasons for which one holds a belief are already in place before one reflects on the normative credentials of one's belief. Kornblith (2012: 21-22) maintains that reflection can be reliable only if we have reliable introspective access to basing relations i.e., the causal history of one's belief.

Consider a case that might help motivate the **Introspection Requirement**. *Bad Detective* believes that the butler committed the murder. Before submitting a report of his investigation to the police, he asks himself, "Am I right in believing the butler is the murderer? Do I really have good reasons to think that?" He introspects on his own reasons for belief. Not long afterwards, he tells himself that he has enough circumstantial evidence to make the case that the butler did it. As it turns out, his false belief that the butler did the murder is *actually based* on racial bias: *Bad Detective* has a deep-seated hatred of people who belong to the same race as the butler. He does not have reliable introspective access to the actual basis for his belief, and when he considers what reasons he has for his belief, he confabulates. While reflecting, *Bad Detective* is unable to see the proper force of the evidence he has because his first-order belief about the butler is maintained due to racial bias. As a result, *Bad Detective* forms the false second-order belief that his first-order belief about the butler is well-proportioned to the evidence. His reflection failed to improve his epistemic condition because he could not accurately identify and correct his own bias.

But do we really need reliable introspective access to basing relations to improve the accuracy of our first-order beliefs via reflection? I think not. Reliable introspective access to basing relations is unnecessary for the project of epistemic amelioration. The causal history of my belief constitutes my *backward-looking* reason for belief. But reflection can proceed in an entirely *forward-looking* manner. To improve my epistemic condition, I do not need to reflect on what my

own reasons are for believing something. Instead, I can pay attention to my evidence and reexamine it carefully. The evidence I have at my disposal gives me *forward-looking* reasons for belief. I need to know what my evidence supports and respond to it accordingly. In doing so, I may get closer to the truth.

To get a sense of how this might work, consider a variant of the case discussed earlier. *Good Detective* believes that the butler committed the murder. Before submitting a report of his investigation to the police, he asks himself, "Did the butler really commit the murder?" He does not explicitly think about what his own reasons are for believing that the butler did the murder. (Let us suppose that like *Bad Detective*, *Good Detective* does not have reliable introspective access to the reasons for his belief. Furthermore, let us suppose that the *Good Detective* would engage in confabulation if he were to seriously reflect on his own reasons for belief). Instead, *Good Detective* considers afresh the complicated body of evidence in his possession. After a careful re-examination of the evidence, *Good Detective* realizes that he has been mistaken all along: the evidence does not point to the butler, it points to the businessman. The businessman had deliberately planted some evidence on the crime scene which was supposed to mislead investigators into thinking that his butler did the crime. Having figured this all out, *Good Detective* changes his mind. He acquires the true belief that the businessman committed the murder. And he does this without making any commitments about the basis on which he formerly believed that the butler committed the murder. *Good Detective* knows what his evidence supports, and he responds to it accordingly. He has improved his epistemic condition. In this way, reflection can proceed in an entirely *forward-looking* manner.[7]

[7] See Mi (2015) for a Confucian way of drawing the forward-looking vs. backward-looking distinction in the context of reflection. Although the distinction Mi draws is similar to my own, his understanding of reflection is not limited to second-order reasoning: he allows that one's reflection could be directed not just at first-order mental states and

I take it that the *Good Detective* case is a kind of possibility proof: *forward-looking* reflection can succeed even if *backward-looking* reflection has failed. In other words, even if one does not have reliable introspective access to basing relations, one can still successfully carry out the project of ameliorating one's first-order beliefs via reflection.

Now, my opponent could concede this point but wonder how often we are in the *good case* as opposed to being in the *bad case*. To reject *The Introspection Argument*, however, the burden is not on me to establish the *positive* claim that *forward-looking* reflection succeeds often enough to count as "reliable" even when *backward-looking* reflection has failed. That said, let me put forward an optimistic, though tentative suggestion. Harman (1986: 41-42) rightly observed that many of our beliefs are formed on the basis of evidence that we have now forgotten. Some of these beliefs are false. If one is to improve upon these false beliefs via reflection, then one must begin by considering afresh the evidence one currently has at one's disposal. Thus, to the extent that we succeed in improving – via reflection – the accuracy of beliefs whose bases we have now forgotten, we do it via *forward-looking* reflection, not *backward-looking* reflection.

Still, one might have positive reasons for thinking that we are more often in the *bad case* rather than the *good case*. One might think that we are biased in all kinds of ways. Plausibly, bias negatively impacts our ability to properly see the force of the evidence, and ultimately leads us to resist epistemically fruitful belief revision. This is what happened with *Bad Detective*.

But are we always, or for the most part, biased? Can we not overcome our biases? These are interesting empirical questions, but they express distinct worries related to *The Efficacy Argument.* That argument attacks the claim that our reflection is causally efficacious in the ways

propositions, but also actions, behaviors, and events (2015: footnote 5). I thank the anonymous reviewer for bringing to my attention Mi's (2015) work on reflection.

we want it to be, which includes overcoming wishful-thinking and bias. I shall critically evaluate *The Efficacy Argument* in section 3.

For now, let me point out that *The Introspection Argument* neither states nor entails the claim that we are always, or for the most part, biased. **Opacity** and **Confabulation** do not – individually or jointly – entail that one will remain unmoved after re-evaluating one's evidence. Nor is it clear that **Opacity** and **Confabulation** increase the likelihood that one is biased. *The Introspection Argument* provides fuel for skepticism regarding the reliability of *backward-looking* reflection i.e., reflection that incorporates introspection on basing relations. It provides no grounds for skepticism regarding the reliability of reasoning in general. For if we were hopelessly biased, all reasoning – either reflective or unreflective – would be unreliable. Surely, that is not what *The Introspection Argument* is supposed to show.

The **Introspection Requirement** stands unmotivated. As illustrated in the *Good Detective* case, we can take steps towards epistemic amelioration so long as we know what our evidence supports. Having reliable access to basing relations is a superfluous requirement on the reliability of reflection. Perhaps the **Introspection Requirement** as formulated is too strong to be plausible. Might there be room for a weaker version of the introspection requirement?

The answer to that question depends on our view of evidence. On a familiar view of evidence, my evidence is constituted only by my mental states. On such a view, even on the *forward-looking* picture of reflection, I would need introspective access to my mental states. For I can know what my evidence supports only if I have access to my evidence. At this point, one might wonder whether the empirical research that Kornblith and Doris cite in support of **Opacity** and **Confabulation** undermines the kind of introspective access we need to get *forward-looking* reflection going. It does not. The relevant empirical research only undermines *backward-looking*

introspection: that is, reliable access to basing relations. Subliminal influences on belief formation and our impressive ability to cook up normative stories to justify our beliefs speak to the fact that we are often in the dark about our reasons for belief. Since one's total evidence is not exhausted by the reasons on which one's beliefs are based, lacking reliable access to basing relations does not entail that we lack reliable access to our evidence *simpliciter.* In the *Good Detective* case, the agent may not have access to the evidence that moved him to believe that 'the butler committed the murder.' But that is no reason to think that he does not have access to other bits of evidence. Thus, even if one could motivate an introspection requirement on the reliability of reflection, we have not been given a plausible argument to think that we would fail to meet such a requirement.

On an alternative view of evidence, one's evidence is not constituted by internal mental states, but by *facts* or *states of affairs*. On this view, *Good Detective*'s revised belief that 'the businessman committed the murder' is justified not on the basis of the beliefs he holds about the businessman's motives and behavior, but on the basis of facts like *the businessman planted misleading evidence*. Of course, on such a view, reflection can proceed in a *forward-looking* manner without the need for any kind of introspective access at all.

Whether or not there is room for some introspection requirement on reflection, the **Introspection Requirement** as stated is too strong to be plausible. If my argument so far is on point, we have earned the right to reject *The Introspection Argument* on the grounds that it imposes a superfluous requirement on the reliability of reflection.

This brings us to *The Efficacy Argument*. This argument concerns whether reflection is causally efficacious i.e., whether it leads us to change our minds for the better. Empirical research – which I have not yet discussed – seems to suggest that reflection is unlikely to prompt a change of mind. *The Efficacy Argument* allows that cases like the *Good Detective* case are possible but

maintains that they are somewhat anomalous. It maintains that reflection, for the most part, fails to increase the accuracy of our first-order beliefs.

3. The Efficacy Argument

The Efficacy Argument is an independent argument against the reliability of reflection. Even if we reflect in an entirely *forward-looking* way, worries concerning the efficacy of such reflection remain. Let us now turn to the argument itself.

The Efficacy Argument

1. **Epistemic Value Reliabilism**: The only determinant of epistemic value is reliability.
2. Therefore, reflection has epistemic value only insofar as it is reliable.
3. **Efficacy Requirement**: Reflection is reliable only if it causally influences one's first-order beliefs.
4. **Anti-Efficacy**: Psychology shows that reflection is for the most part causally inert, or that it influences our first-order beliefs in *the wrong kind of way*.
5. Therefore, at worst, reflection is unreliable, and at best, it leaves our first-order beliefs in no better epistemic shape than they were before.
6. Therefore, reflection has little or no epistemic value.

The argument shares premises 1-2 with *The Introspection Argument.* The **Efficacy Requirement** is eminently plausible. **Anti-Efficacy** states that we often fail to meet the **Efficacy Requirement**.

Clearly, the crucial premise here is **Anti-Efficacy**, in support of which philosophers like Kornblith (2012: 20-26) and Doris (2015: 92-97) have marshalled a range of empirical studies. This empirical research is vast. Next, I provide a brief summary of the main findings of this research.

3.1. The Empirical Evidence for Anti-Efficacy

Studies from empirical psychology provide evidence not just for **Opacity** and **Confabulation**, but also for the following phenomenon:

> **Epiphenomenality**: When we reflect on the normative adequacy of the reasons for our beliefs, our higher-order deliberation is often either:
>
> (a) causally inert: it does not influence *why* we continue holding onto a belief (**Inertness**), or
>
> (b) it causally influences our first-order beliefs in *the wrong kind of way* (**Counterproductivity**).

Note that **Opacity** and **Confabulation** do not individually or jointly entail **Epiphenomenality**. It is one thing to say that people's higher-order deliberation is often disconnected from their actual reasons for belief or action, and quite another to say that such higher-order deliberation does not causally alter one's first-order mental states *after the fact*. **Epiphenomenality** is supported by studies on motivated cognition, self-enhancement, and confirmation bias.

Motivated cognition and self-enhancement, as the names suggest, involve a tendency in individuals to think highly of their own credentials (epistemic or otherwise). Strictly speaking, motivated cognition refers to believing what one wants to believe. However, almost everyone wants to believe better of themselves. People find ingenious ways of making their doxastic and

conative lives "look good."[8] Such factors strongly dispose people in preserving the first-order beliefs they have already formed. Thus, the evidence one has at one's own disposal is taken at face value or selectively scrutinized. Evidence in favor of one's beliefs is better remembered. New evidence supporting one's beliefs is easily accepted, whereas opposing evidence is more easily rejected. Moreover, the process of searching for new evidence is skewed in such a way that the results are likely to favor what one already believes.[9]

Kornblith's (2012: 25-26) overall assessment of this research is as follows. We have little reason to hope that episodes of reflection alter, let alone improve, our epistemic condition. Reflection suffers from **Inertness**: episodes of higher-order deliberation on one's reasons for belief are little more than exercises in "self-congratulation." Instead of resulting in epistemically fruitful belief revision, they produce an agent who has a false sense of security about the accuracy of their first-order beliefs. If anything, the agent becomes more entrenched and dogmatic after subjecting their beliefs to critical scrutiny. Thus, reflection is also susceptible to **Counterproductivity**: even when reflection exerts causal influence on one's first-order beliefs, it often does so in the *wrong kind of way*.

For the sake of argument, I shall grant Kornblith's assessment of the empirical research, but with one important qualification. The empirical research – even if it is shown to be solid – only impugns the reliability of individual reflection. However, for *The Efficacy Argument* to go through, **Anti-Efficacy** must be true for reflection *in general*.

In what follows, I will argue that the case for **Anti-Efficacy** has not been made for reflection *in general*. To be clear, I will not be arguing for the *positive* claim that social reflection

[8] On self-enhancement, and motivated cognition more generally, see: Kunda (1990), Alicke et al. (2001: 9), papers in Gilovich et al. (2002: Part I, Sections C and D), Dunning (2006), and Dufner et al. (2012: 538).
[9] On our tendency to preserve what we already believe, and confirmation bias more generally see: Fyock and Stangor (1994), Nickerson (1998), and Taber and Lodge (2006).

is reliable. My goal is to rebut the *negative* claim that social reflection is bound to be unreliable for much the same reasons that putatively undermine the reliability of individual reflection. Since my goal is not to provide a positive argument for the reliability of social reflection, I will only be providing a provisional defense of the *comparative* claim that social reflection does better than individual reflection in several respects, rather than the *non-comparative* claim that social reflection does well enough overall to count as reliable.

I will argue that – under certain conditions – social reflection is not plagued by **Inertness** or **Counterproductivity**. In other words, it is not self-congratulatory, it does not make us more intransigent, and it does not lull us into a false sense of epistemic security.

3.2. Resisting Anti-Efficacy in the Context of Social Reflection

To see why worries about the putative **Epiphenomenality** of individual reflection do not generalize to social reflection, we must focus on certain unique features of social reflection.

First, social reflection often involves dealing with challenges to our beliefs in a way that individual reflection does not. Second, challenges to our beliefs encountered socially are harder to push under the rug, so to speak. Consequently, when faced with such challenges – as opposed to challenges encountered individually – one is more likely to change one's mind. Third, due to what I will call the *critical function* of social reflection, one's higher-order deliberation on normative reasons for belief is far from "self-congratulatory." Instead, one is confronted with powerful objections and arguments against one's beliefs. Fourth, under a range of circumstances, one can harness the *critical function* of social reflection to bring about a change of mind for the better.

The first two features speak against the **Inertness** of social reflection, whereas the last two raise doubts about its **Counterproductivity**. Let us consider each pair of features in turn.

3.2.1. Resisting Inertness

The evidence for **Epiphenomenality** – which I have granted for the sake of argument – suggests that we do not often reflect individually on the evidence we have for our beliefs. If we are strongly predisposed to preserve the first-order beliefs we have already formed, then it is hardly surprising that we do not investigate the accuracy of our first-order beliefs by critically reflecting on the evidence we have for them. Furthermore, even when we do manage to reflect on the evidence we have for our beliefs, we rarely disagree with ourselves. More often than not, one's individual reflection regarding the normative question of whether one should believe p will be in harmony with one's first-order belief that p – even if one entertains objections or considers contrary evidence, one will, due to confirmation bias and the like, eventually end up "rationalizing" what one already believes (Evans and Wason 1974).

So far, I have mentioned contingent limitations of individual reflection. There are normative limitations as well. It is widely assumed that rationality demands that an agent's first-order beliefs be consistent with her higher-order beliefs. An agent who believes that p, and that 'I do not really have good evidence for my belief that p' is being epistemically irrational.[10] Relatedly, there is rational pressure on us to not disagree with ourselves. An agent who believes the conjunction 'p but after carefully reflecting on my evidence, ~p', is epistemically irrational. The norms of rationality therefore prohibit disagreeing with ourselves.

It seems, therefore, that when we form false first-order beliefs, we dig ourselves into a hole too deep for individual reflection to succeed in correcting our false beliefs. Whether or not this really is the case with individual reflection, this need not be the case with social reflection.

[10] The view that epistemic akrasia is irrational is not, however, uncontroversial. See Horowitz (2022) for arguments the view that epistemic akrasia is not irrational.

For starters, *interpersonal* disagreement is more common than *intrapersonal* disagreement. When I assert my belief that p, my interlocuter might challenge my assertion and I will be invited to consider whether p, or why p. When I *publicly* – as opposed to *privately* – assert my belief that p, I am more likely to encounter disagreement. This fact is both normatively and empirically significant.

Disagreement with others puts rational pressure on me to reconsider whether I have responded appropriately to my evidence. The extent of rational pressure to reconsider my response may vary, depending on the nature of the disagreement. If my interlocutor and I have the same evidence, then one of us has responded wrongly to the evidence, and it would be dogmatic to discount the possibility that I am the guilty party. If my interlocutor has different evidence, then one of us has better evidence, and it would be dogmatic to discount the possibility that mine is inferior. The mere fact of disagreement raises the possibility that I might be wrong. In a social context where the truth of p is being challenged, there is *some* rational pressure on me to reconsider whether p. In this way, public disagreement is normatively significant.

Public disagreement is also empirically significant. It often triggers social reflection. When agents disagree about some matter, they often pursue their disagreement collaboratively, by jointly pursuing the question, "What does the evidence give us good reason to believe?" More importantly, social reflection that stems from interpersonal disagreement is more likely to result in a change of mind, in comparison with individual reflection.[11] As Mercier (2020: 47-50) points out, with respect to "non-sensitive" cases of disagreement (e.g., arithmetic, geography etc.), interpersonal disagreement usually results in a change of mind. With "sensitive" (or emotionally

[11] For discussion on this comparative point, see Mercier and Sperber (2017: 9-10, 233-235, 247-250, 264-265, 295).

fraught) topics like religion, politics, or morality, we sometimes witness the so-called "backfire effect": presented with contrary opinions or arguments from interlocutors, individuals become more entrenched in their views (Nyhan and Reifer 2010). As we shall see later, this need not rule out the prospects of social reflection in increasing the reliability of our beliefs regarding "sensitive" topics.

To recap so far: (1) we are more likely to encounter disagreement when we reflect socially as opposed to reflecting individually, (2) disagreement puts rational pressure on us to reassess our responses to our evidence, (3) encountering disagreement publicly is more causally efficacious than intrapersonal disagreement, and (4) likely effects of encountering disagreement publicly include a change of mind. So far, I have said nothing to establish that social reflection increases the reliability of our first-order beliefs. When we reflect socially, are we more likely to change our minds for the better? If so, how? This brings us to **Counterproductivity.** Part of the answer to those questions lies in the *critical function* of social reflection, which speaks to the potential of social reflection in generating a better quality of higher-order deliberation on normative reasons for belief.

3.2.2. Resisting Counterproductivity

For the most part, we are lousy at coming up with good arguments against our own beliefs. However, we are much better when it comes to finding good objections to, and producing solid arguments against, others' beliefs. This is what I shall call the *critical function* of social reflection. Psychology tells us that we are more critical of others' professed beliefs, and that we are quicker and better at undermining them.

A study done by Kuhn et al. (1994) asked mock jurors to give their verdict on a particular case. Most mock jurors, when presented with an alternative verdict, were able to produce a counterargument to it. Another study done by Shaw (1996) showed that all the participants in the study could quickly come up with counterarguments against a claim which was not theirs. In a study done by Trouche et al. (2016) demonstrating the "selective laziness" of human reasoning, researchers attempted to trick participants into thinking that their own arguments for a claim were someone else's. Over half the participants who fell for the trick rejected the argument (which was in fact theirs). Furthermore, participants were more likely to reject invalid arguments, which suggests that they were better able to tell valid from invalid arguments when they perceived the arguments as not theirs. Resnick et. al. (1993) created groups of three participants who disagreed over an issue. In such an "argumentative setting," what they observed was that participants demonstrated skill in reasoning, including the ability to understand argument structure, to build complex arguments of their own, and to identify and undermine premises of others' arguments. The key point here is that the critical function of our reasoning faculties is enhanced in a social, argumentative setting.[12]

In the game of giving and asking for reasons, the listener remains "epistemically vigilant": more often than not, instead of immediately accepting the speaker's assertion as true, the listener asks clarificatory questions, and offers pushback (Sperber et. al. 2010). Speakers are typically aware of this, and so they tailor-make their arguments to convince their listeners, say, by appealing to the listener's own background beliefs. As Mercier and Sperber (2017: 222-237) have shown elsewhere, listeners' "epistemic vigilance" acts as quality control viz-a-vis the speakers' claims and arguments. The speaker's "myside bias" or motivated reasoning is mitigated: although the

[12] For a review of the relevant empirical evidence, see Mercier and Sperber (2011: 61-63).

speaker retains a strong tendency to justify her own beliefs, she is forced to produce contextually persuasive arguments for them. As Mercier and Sperber (2011: 60) put it, this speaker-listener dynamic, among other things, makes communication (on the whole) more reliable and hence more advantageous.

Thus, in addition to factors (1)-(4) discussed earlier, (5) the critical function of social reflection, and (6) the dynamic between speakers and listeners in the game of giving and asking for reasons, all speak in favor of the potential of social reflection in generating a better quality of higher-order deliberation on normative reasons for belief. Furthermore, even if listeners are on guard, disagreement often results in a lowering of confidence in their views, which sometimes culminates in a change of mind.

Now, a crucial question is: under what conditions do individuals harness the potential of social reflection to successfully ameliorate their own epistemic condition? In response, I will discuss three conditions that favor epistemic amelioration in the context of social reflection. These are: group settings, viewpoint diversity, and the formal division of cognitive labor. Let us take each in turn.

Group Settings. Empirical research suggests that in groups of three or more, individuals are better at solving "logical" or "intellective" problems i.e., those that have objective answers. (As we shall see, group settings remain useful in tackling other kinds of problems as well).

Take for instance the famous Cognitive Reflection Test (CRT). The CRT, among other things, is a tool to measure people's tendency to favor intuitive or gut reactions over individual reflection (Frederick 2005). One of the questions on this test goes as follows: if it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? Most

people answer 100 minutes. The correct answer is 5 minutes. McRaney (2022: Ch. 7) reports that reasoning alone, 83 percent of people who take this test under laboratory conditions get at least one among three such questions wrong, and about 1/3rd get all three questions wrong. In groups of three or more, however, no participant gets any question wrong. Mercier and Sperber (2011: 62-63) report that participants taking the famous Wason selection task (a logic puzzle which tests deductive reasoning skills), when reflecting individually, perform very poorly. Performing in a team, however, drastically improves individual performance on the test. In these studies, at least one participant sees the correct answer, and the ensuing debate sees other participants change their minds for the better. As McRaney puts it, the path to epistemic amelioration goes like this: "lazy reasoning, disagreement, evaluation, argumentation, and truth." (2022: 197)

One might object that the success of social reflection seems to be parasitic on the success of individual reflection: at least one instance of individual reflection must lead to epistemic amelioration for other individuals in the group to be better off epistemically. The matter is complicated. In many cases, no single individual has the correct answer, and others might be partly wrong or partly right. But the group collectively manages to converge on the right answer (Laughlin et. al 2006). Evidence for this phenomenon can be gathered from developmental psychology as well (Mercier 2011).

Viewpoint Diversity. Strength in numbers might be sufficient to bolster individual performance when it comes to mathematics and logic. In other domains, while numbers remain important, what is even more important in getting closer to the truth is viewpoint diversity.

Consider two kinds of viewpoint diversity: *cognitive diversity* and *ideological diversity*. Cognitive diversity concerns the diverse ways in which people approach (diversity of perspective), conceptualize (diversity of interpretation), and tackle a problem or question (diversity of

heuristics). Cognitive diversity, resulting from a conglomeration of randomly selected individuals, is good for problem-solving in general (Hong and Page 2004). The Diversity Trumps Ability Theorem states that, given certain conditions, "a randomly selected collection of problem solvers almost always outperforms a collection of the best individual problem solvers" (Page 2007).

Landemore (2012: Ch. 4) discusses several real-life examples of political problem-solving which vindicates this research. Consider one such case. A New Haven neighborhood had a terrible mugging problem. Stakeholders and local authorities held meetings to deliberate on possible solutions. Those deliberating were a cognitively diverse bunch: regular citizens, the police, engineers, accountants. Over the course of the discussions, the deliberators moved away from the most sub-optimal solution (the police car posted at the corner of the dangerous block) to the more compelling one (solar lamps on the bridge). The solutions offered by the police – the experts in this kind of situation – were confrontational in nature. It took a motley of perspectives to break away from one-dimensional thinking. The solution, since its implementation in 2010, has had remarkably success. This same story could be told in terms of epistemic amelioration: deliberators' first-order beliefs about the optimal way to curb mugging in their neighborhood changed for the better i.e., went from being less accurate to more accurate. Individual stakeholders, including the experts, were probably worse off deliberating about the optimal solution on their own. With numbers, and cognitive diversity, the deliberators got closer and closer to the truth. This is an example of reflection increasing reliability in the domain of practical rationality i.e., questions concerning the best means of achieving a certain end.

Reflection can increase reliability not just on matters concerning the desirability of certain means, but also the desirability of the ends themselves. Ideological diversity is key to increasing reliability, or making progress, in moral and political domains. Ideological diversity refers to a

diversity of values (moral, political, aesthetic) and worldviews (comprising various metaphysical, epistemological beliefs) among people. Gaus (2016: 230) argues that along with cognitive diversity, ideological diversity is the "source of improvements in the moral constitution of the Open Society," and that truths about justice are best sought under an arrangement where communities who disagree on what a just society looks like are able to live according to their ideals. Communities learn from their "experiments of living" – to use Mill's (1859/2009: 95) phrase – and share their insights with others.

Ideological diversity is at the heart of much political disagreement. Examples from history testify that such disagreement, pursued through reasoned debate, can be a major factor in bringing about moral and political progress. Drescher (2009: 205-241) tells us how the abolitionist movement of the 18th and 19th centuries, which sought to end slavery and slave-trade in British colonies, achieved its end through vigorous parliamentary debates, the publication of pamphlets and newspaper articles, and public meetings. Over the course of decades, members of the British parliament as well as the general public, including vociferous anti-abolitionists, became convinced of the evil of slavery, and the need to do something about it. Likewise, in colonial India of the early 20th century, prominent Indian leaders publicly debated the practice of untouchability. Arriving at a consensus through reasoned debate, a political ban on the practice of untouchability became part of the mainstream political agenda in India. Soon after India achieved independence from Britain in 1947, this agenda was implemented. Untouchability was criminalized (Galanter 1969). In some cases, however, social reflection fails to bring about a change in the attitudes of people. Vigorous parliamentary debates in the United States failed to bring any similar laudatory results, and the issue of slavery led to civil war (Lowance Jr. 2003).

While it is true that strength in numbers and ideological diversity are not sufficient to precipitate moral and political progress, we cannot ignore the overall impact of social reflection on societies across the world, especially in the last three centuries or so. In Western societies, the latter half of the 18th century brought what Pinker (2011: 133) calls the "Humanitarian Revolution." The revolution gave rise to the concept of human rights and the ideology of humanism, which were deployed to minimize or abolish institutionalized violence and human suffering more generally. Moral agitators, and the public debates they began, influenced public legislators to abolish immoral practices like blood sports, public hangings, cruel punishments, debtors' prisons, etc. Such transformations in society often go hand in hand with a change in people's sensibilities i.e., a change in their moral and political beliefs (2011: 168-169). Closer to our century, in several democracies which allow the free expression of ideas, similar processes have contributed to the decriminalization of homosexuality, the empowerment of women and vulnerable minorities, and much else (Mansbridge 1999).

All that said, in group settings, too much ideological diversity and disagreement might be a problem. De Ridder (2022: 226-227) argues that "deep disagreements" do not foster epistemic amelioration because they cause citizens to see each other as "less than fully rational, as morally subpar, or worse." Deliberation which features "deep disagreement" is likely to trigger the "backfire effect" i.e., when people become more entrenched in their views. On the other hand, Esterling et al. (2015) report that "moderate ideological difference" is good for democratic deliberation. Moderate ideological difference and cognitive diversity might be the keys to overcoming the "backfire effect."

Formal Division of Cognitive Labor. When it comes to questions in mathematics, natural science, technology, and theoretical philosophy, ideological diversity might not be as important.

Epistemic amelioration can be accelerated when communities institute formal mechanisms to tap into, and manage, cognitive diversity. The instances of social reflection that I have been discussing till now are *informal* or *semi-formal* e.g., discussion and debate among individuals, in civil society at large, in controlled experimental settings, and so on. Social reflection is *formal* when it is more structured and is subject to more rigorous standards. Thus, a graduate-level seminar, or an academic conference on topic X are formal counterparts of a discussion on X between laypeople at a coffee shop. Formalized social reflection is not restricted to the spoken word. Peer-reviewed academic journals, books, and symposia are powerful ways in which individuals get critical feedback on their views. Broadly speaking, laboratory settings, workshops, degree programs, research institutes etc. all facilitate social reflection.

Consider scientific communities. On Bird's (2010) picture, scientific communities are groups that are bound together by a mutual interdependence brought about via a division of cognitive labor. "Social cognitive structures" like scientific communities (a) have characteristic outputs that are propositional in nature (propositionality), (b) have characteristic mechanisms whose function is to ensure or promote the chances that the outputs in (a) are true (truth-filtering), and (c) are so organized that the outputs in (a) are accessible to individual members who need it (accessibility). An instance of (a) is a journal article; an instance of (b) is the peer-review process; and an instance of (c) is the publication of the article in print and online. That by being part of scientific communities, individual members produce scientific knowledge is a fact well appreciated. What is perhaps less appreciated is a process that naturally runs parallel to scientific knowledge production. In producing scientific knowledge, scientists' first-order beliefs about their research topic can change for the better.

Now, one might worry whether social reflection has any utility in disciplines like theoretical philosophy. After all, philosophers sometimes look like paradigm cases of entrenched disagreement (Chalmers 2015). If philosophers have consistently failed to converge on the truth, then it seems that any kind of reflection (individual or social) on the subject matter of philosophy cannot be reliable. Comparing the situation of philosophy with science may make it seem that it is the subject matter that is more relevant to the question of whether we can be reliable with respect to a domain, and not how we reflect on the subject matter.[13]

I do not think these grounds warrant either (i) mitigated skepticism about the utility of social reflection in science, or (ii) blanket skepticism of its utility in theoretical philosophy. Plausibly, formally structured social reflection partly explains our reliability with respect to the subject matter of science. The division of scientists into research teams with different interests and motivations is ultimately beneficial for science (Kitcher 1990). Such a setup increases the likelihood of a thorough investigation of a scientific problem, with each research team presenting the best arguments for their hypothesis, and subjecting rival hypotheses to maximum scrutiny. The importance of formally structured social reflection in science becomes evident when we compare, say, the fecundity of modern physics with pre-scientific investigations into the physical world done in ancient and medieval times. On the other hand, it is true that we are not equally reliable in philosophy, even though the structures of social reflection in philosophy are similar to those in science. What explains the difference is not any deficiency in social reflection, but the fact that philosophy is an exception – there are unique features of philosophy which make progress harder to come by e.g., it is hard to measure progress in philosophy.

[13] I would like to thank an anonymous reviewer for pressing this worry.

One could appreciate the force of these points while maintaining the weaker, *comparative* claim that social reflection does better than individual reflection in making us more reliable in philosophy. Even if convergence to the truth on the "big questions" remains unlikely, philosophers have converged to the truth on philosophy's "smaller questions." As Chalmers (2015: 16) puts it, analytic philosophers have knowledge of various "negative and conditional theses, of frameworks available to answer questions, of connections between ideas, of the way that arguments bear for or against conclusions, and so on." Taking a closer look at some of these cases reveals that it is social reflection, as opposed to individual reflection, that is doing most of the explanatory work. For instance, due to the debate on knowledge provoked by Gettier (1963), most philosophers now subscribe to the negative thesis that whatever else knowledge might be, it is *not merely* justified true belief. Moreover, most philosophers now have a better understanding of the constraints on knowledge. And this convergence can be explained in large part due to mechanisms of formally structured social reflection in analytic philosophy – namely, graduate seminars, conferences, and peer-review.

3.3. Summary and Upshot

To sum up: empirical evidence suggests that social reflection – under certain conditions – is not beset with the same distortions that plague individual reflection. Due to interpersonal disagreement, which often spurs social reflection, an individual is more likely to change their mind. The critical function of social reflection is enhanced in group settings such as public debates, discussions where cognitively diverse individuals try to solve problems, public fora with moderate ideological diversity, or via the peer-review process. In such social contexts, one's chances of changing one's mind for the better increase. Thus, in a range of conditions, social reflection fares

better than individual reflection in increasing the accuracy of our first-order beliefs. And so, the case for the unreliability of individual reflection does not generalize to social reflection.

Now, one could still wonder about the extent to which social reflection is carried out under the kinds of favorable conditions I mentioned. This remains an important empirical issue in adjudicating the overall reliability of social reflection – a task which I have not undertaken in this paper for two reasons. First, the extent to which social reflection is carried out under favorable conditions is an issue that cannot be adjudicated without a careful evaluation of the distinct *macro-level* distortions faced by social reflection. While I briefly discuss these distortions in the conclusion, such an evaluation is beyond the scope of this paper. Second, given my argumentative aims, this evaluation is unnecessary. To defend the reliability of reflection against Kornblith's (2012) critique, as opposed to the ambitious task of providing a *positive* argument for the overall reliability of reflection, it is enough to show that the case against the reliability of individual reflection does not generalize to social reflection. My discussion on the unique features of social reflection suffices for that purpose. Furthermore, it gives us *prima facie* grounds for optimism regarding the reliability of social reflection.

I therefore reject *The Efficacy Argument* on the grounds that we lack sufficient evidence to believe that **Anti-Efficacy** is true for reflection in general.

4. Conclusion

Defenders of reflection are often accused of having an overly individualistic focus. As it turns out, this charge applies to some critics of reflection. The selective focus on Cartesian conceptions of reflection is unjustified because it misrepresents a long-standing tradition of philosophical thinking

on reflection. Consider two figures from this tradition: Socrates and Mill. The Socratic method – the *elenchus* – is an argumentative dialogue between individuals, with the aim of gaining a more refined understanding of important philosophical concepts through collaboration. And social reflection is at the heart of Mill's *On Liberty*. One of Mill's strategies to defend the free expression of ideas is to highlight the epistemic payoffs of public debate for individuals and society. Mill's (2009: 63) remarks on the importance of critical reflection must be understood in that context.

Although social reflection is not equally plagued by the distortions faced by individual reflection, like any other naturally or socially evolved mechanism, it has distortions of its own. In group settings such as public discussions, if there are too many like-minded people, we witness belief polarization: people's views on a given topic become more extreme and partisan (Myers and Bach 1974). Ideological conformity promotes "groupthink," which often runs contrary to the project of epistemic amelioration (Leanna 1985). If we find ourselves in "epistemic bubbles," or in "echo chambers" wherein evidence is selectively filtered and views conforming to only one kind of ideology are constantly presented as attractive, then social reflection is very likely to be unreliable (Nguyen 2020). These scenarios present a new threat to the epistemic value of social reflection. How widespread are they? And can they be the basis of a new empirical critique of the epistemic value of reflection in general?

Mercier (2020: 211-214) reviews a range of empirical studies and concludes that the situation on the ground is not as bad as it seems. Here, I briefly report the conclusions of some of these studies. The degree of political polarization in the United States is often exaggerated. The percentage of independents (people who are neither Republicans nor Democrats) has increased in recent years. Most Americans think of themselves as moderates, rather than conservatives or liberals, and have done so for the last forty years or so. Even on social media, reports of

polarization are exaggerated. On Twitter, while 1% of the most active users behave according to polarization narratives, individuals in the other 99% share more moderate content than they receive. Some research suggests that the idea that we are trapped in echo chambers is a bigger myth than the idea that polarization is dangerously on the rise. For instance, in Germany and Spain, studies show that social media users are embedded in ideologically diverse networks, and that only 8% of online adults in the United Kingdom are at risk of being confined to an echo chamber. Social media outlets may not be as pernicious as they seem. Compared to a group that stopped using Facebook for a month, the group that kept using Facebook did not develop more polarized attitudes. Another study shows that Facebook usage contributed to depolarization, due to exposure to dissenting views.

While this research is encouraging, a thorough evaluation of the overall reliability of social reflection remains a topic for future work. In this paper, I have endeavored to show that the recent critique of the epistemic value of reflection, though insightful in some ways, is limited in scope. Even if Kornblith's (2012) critique undermines the reliability of individual reflection, it does not thereby undermine the reliability of social reflection. We do not yet have a successful argument against the epistemic value of reflection in general.

Acknowledgements

I am especially grateful to Declan Smithies, Tristram McPherson, and Abraham Roth for detailed feedback on earlier drafts of this paper. I would also like to thank Hilary Kornblith, Eden Lin, Daniel Buckley, Seungsoo Lee, Vaughn Papenhausen, Nathan Dowell, and Preston Lennon for valuable discussion. Finally, I'd like to say a word of thanks to two anonymous *Ergo* referees for their thoughtful comments and suggestions.

References

Alicke, Mark D., Vredenburg, Debbie. S., Hiatt, Matthew, and Govorun, Olesya (2001). The Better than Myself Effect. *Motivation and Emotion*, 25, 7-22.

Alston, William (1989). *Epistemic Justification: Essays in the Theory of Knowledge*. Cornell University Press.

Banaji, Mahzarin R. and Greenwald. Anthony G. (2013). *Blindspots: Hidden Biases of Good People.* Delacorte Press.

Bertrand, Marianne and Mullainathan, Sendhil (2003). "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." MIT Department of Economics Working Paper No. 03-22. Available at SSRN: http://ssrn.com/abstract=422902 or DOI: DOI: 10.2139/ssrn.422902

Bonjour, Lawrence (1985). *The Structure of Empirical Knowledge.* Harvard University Press.

Brandom, Robert (1994). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press.

Bird, Alexander (2010). Social knowing: The Social Sense of 'Scientific Knowledge'. *Philosophical Perspectives*, 24(1), 23-56.

Burge, Tyler (1996). Our Entitlement to Self-Knowledge. *Proceedings of the Aristotelian Society*, 96(1), 91-116.

Carruthers, Peter (2009). How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition. *Behavioral and Brain Sciences*, 32, 121-38.

Chalmers, David (2015). Why Isn't There More Progress in Philosophy? *Philosophy*, 90(1), 3-31.

Clifford, William K. (1999). The Ethics of Belief. In Timothy Madigan (Ed.), *The ethics of belief and other essays* (70-96). Prometheus. (Original work published 1877).

De Ridder, Jeroen (2022). Deep Disagreements and Political Polarization. In Elizabeth Edenberg and Michael Hannon (Eds.), *Political Epistemology* (226-44). Oxford University Press.

Descartes, Rene (1990). *Meditations on First Philosophy*. (George Heffernan, Trans.). University of Notre Dame Press. (Original work published 1641).

Doris, John M. (2015). *Talking to Ourselves: Reflection, Ignorance, and Agency.* Oxford University Press.

Dufner, Michael, Denissen, Jaap J., Van Zalk, Maarten, Matthes, Benjamin, Meeus, Wim H., Van Aken, Marcel A., and Sedikides, Constantine (2012). Positive Intelligence Illusions: On the Relation Between Intellectual Self-Enhancement and Psychological Adjustment. *Journal of Personality*, 80, 537-72.

Dunning, David (2006). *Self-Insight: Roadblocks and Detours on the Path to Knowing Thyself.* Psychology Press.

Drescher, Seymour (2009). *Abolition: A History of Slavery and Antislavery*. Cambridge University Press.

Eskine, Kendall. J., Kacinik, Natalie A., and Prinz, Jesse J. (2011). A Bad Taste in the Mouth: Gustatory Disgust Influences Moral Judgment. *Psychological Science*, 22(3) 295-9.

Estabrooks. George H. (1957). *Hypnotism (Revised Edition)*. E. P. Dutton.

Esterling, Kevin. M., Fung, Archon, and Lee, Taeku. (2015). How Much Disagreement is Good for Democratic Deliberation? *Political Communication*, 32(4), 529-51.

Evans, Jonathan and Wason, Peter. (1974). Dual Processes in Reasoning? *Cognition*, 3(2), 141-54.

Frankfurt, Harry (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy*, 68(1), 5-20.

Frederick, Shane (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25-42.

Fyock, Jack and Stangor, Charles (1994). The Role of Memory Biases in Stereotype Maintenance. *British Journal of Social Psychology,* 33(3), 331–43.

Harman, Gilbert (1986). *Change in View: Principles of Reasoning.* MIT Press.

Galanter, Marc (1969). Untouchability and the Law. *Economic and Political Weekly,* 4(1/2), 159-70.

Gardner, Wendi L., Gabriel, Shira., and Lee, Angela Y. (1999). "I" Value Freedom But "We" Value Relationships: Self-Construal Priming Mirrors Cultural Differences in Judgment. *Psychological Science*, 10, 321-6.

Gaus, Gerald (2016). *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton University Press.

Gazzaniga, Michael S. (2000). Cerebral Specialization and Interhemispheric Communication: Does the Corpus Callosum Enable the Human Condition? *Brain*, 123, 1293-1326.

Gettier, Edmund (1963). Is Justified True Belief Knowledge? *Analysis*, 23(6), 121-23.

Gilovich, Thomas, Griffin, Dale, and Kahneman, Daniel (Eds.) (2002). *Heuristics and Biases: The Psychology of Human Judgment*. Cambridge University Press.

Haley, Kevin and Fessler, Daniel (2005). Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game. *Evolution and Human Behavior*, 26, 245-256.

Hirstein, William. (2005). *Brain Fiction: Self-Deception and the Riddle of Confabulation.* MIT Press.

Hong, Lu and Page, Scott (2004). Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46), 16385-89.

Horowitz, Sophie (2022). Higher-Order Evidence. *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/fall2022/entries/higher-order-evidence/>.

Joshi, Hrishikesh (2021). *Why it's OK to Speak Your Mind*. Routledge.

Kahneman, Daniel and Tversky, Amos (1982). The Simulation Heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.

Kelly, Daniel. (2011). *Yuck!: The Nature and Moral Significance of Disgust.* MIT Press.

Kitcher, Philip (1990). The Division of Cognitive Labor. *Journal of Philosophy*, 87(1), 5-22.

Kornblith, Hilary (2012). *On Reflection.* Oxford University Press.

Korsgaard, Christine (1996). *The Sources of Normativity.* Cambridge University Press.

Kuhn, Deanna, Weinstock, Michael, and Flaton, Robin (1994). How well do jurors reason? Competence dimensions of individual variation in a juror reasoning task. *Psychological Science*, 5(5), 289–96.

Kunda, Ziva (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, 108, 480-98.

Landemore, Hélène (2012). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton University Press.

Latane, Bibb and Darley, John. (1970). *The Unresponsive Bystander: Why Doesn't He Help?* Appleton-Century-Crofts.

Laughlin, Patrick R., Hatch, Erin C., Silver, Jonathan S., and Boh, Lee (2006). Groups Perform Better than the Best Individuals on Letters-to-Numbers Problems: Effects of group size. *Journal of Personality and Social Psychology*, 90(4), 644–51.

Leana, Carrie R. (1985). A Partial Test of Janis' Groupthink Model: Effects of Group Cohesiveness and Leader Behavior on Defective Decision Making. *Journal of Management*, 11(1), 5–17.

Lowance Jr., Mason I. (Ed.) (2003). *A House Divided: The Antebellum Slavery Debates in America, 1776-1865.* Princeton University Press.

Leite, Adam (2004). On Justifying and Being Justified. *Philosophical Issues*, 14(1), 219-53.

Maier, Norman R. F. (1931). Reasoning in Humans. II. The Solution of a Problem and its Appearance in Consciousness. *Journal of Comparative Psychology*, 12, 181-94.

Mansbridge, Jane (1999). Everyday Talk in the Deliberative System. In Stephen Macedo (Ed.), *Deliberative Politics: Essays on Democracy and Disagreement* (211-39). Oxford University Press.

McRaney, David (2022). *How Minds Change: The Surprising Science of Belief, Opinion, and Persuasion*. Portfolio, Penguin.

Mercier, Hugo (2011). Reasoning Serves Argumentation in Children. *Cognitive Development*, 26 (3), 177-191.

Mercier, Hugo (2020). *Not Born Yesterday: The Science of Who We Trust and What We Believe.* Princeton University Press.

Mercier, Hugo and Sperber, Dan (2010). "Epistemic Vigilance." *Mind & Language*, 25(4), 359-93.

Mercier, Hugo and Sperber, Dan (2011). "Why Do Humans Reason? Arguments for an Argumentative Theory" *Behavioral & Brain Sciences,* 34, 57-111.

Mercier, Hugo and Sperber, Dan (2017). *The Enigma of Reason*. Harvard University Press.

Meritt, Melissa (2018). *Kant on Reflection and Virtue*. Cambridge University Press.

Mi, Chienkuo. (2015). What is Knowledge? When Confucius Meets Ernest Sosa. *Dao: A Journal of Comparative Philosophy*, 14(3), 355-67.

Mill, John S. (2009). *On Liberty.* The Floating Press. (Original work published 1859).

Moran, Richard (2001). *Authority and Estrangement*. Princeton University Press.

Myers, David G., & Bach, Paul J. (1974). Discussion Effects on Militarism-Pacifism: A Test of the Group Polarization Hypothesis. *Journal of Personality and Social Psychology*, 30(6), 741–47.

Nguyen, C. Thi (2020). Echo Chambers and Epistemic Bubbles. *Episteme*, 17(2), 141-61.

Nickerson, Raymond S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175–220.

Nietzsche, Friedrich (1974). *The Gay Science*. (Walter Kaufmann, Trans.). Vintage Books. (Original work published 1882).

Nisbett, Richard E., & Wilson, Timothy D. (1977). Telling More than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, 84(3), 231–59.

Nyhan, Brendan and Reifler, Jason (2010). When Corrections Fail: The Persistence of Political Misperceptions. *Political Behavior*, 32: 303-30.

Page, Scott (2007). Making the Difference: Applying a Logic of Diversity. *Academy of Management Perspectives*, 21(4), 6-20.

Peacocke, Christopher (1996). Our Entitlement to Self-Knowledge: Entitlement, Self-Knowledge, and Conceptual Redeployment. *Proceedings of the Aristotelian Society*, 96(1), 117-58.

Pettit, Philip (2007). *Rationality, Reasoning and Group Agency. Dialectica*, 61(4), 495-519.

Pinker, Steven (2011). *The Better Angels of Our Nature: Why Violence Has Declined.* Viking Penguin.

Resnick, Lauren B., Salmon, Merrilee, Zeitz, Colleen M., Wathen, Sheila H., and Holowchak, Mark (1993). Reasoning in Conversation. *Cognition and Instruction*, 11(3–4), 347–64.

Shah, Nishi and Velleman, J. David (2005). Doxastic Deliberation. *Philosophical Review*, 144(4), 497-534.

Shaw, Victoria F. 1996. The Cognitive Processes in Informal Reasoning. *Thinking & Reasoning*, 2, 51–80.

Smithies, Declan (2015). Why Justification Matters. In J. Greco, & D. Henderson (Eds.), *Epistemic Evaluation: Point and Purpose in Epistemology* (224-44). Oxford University Press.

Smithies, Declan (2019). *The Epistemic Role of Consciousness*. Oxford University Press.

Taber, Charles S. and Lodge, Milton (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science*, 50(3), 755-69.

Trouche, Emmanuel, Johansson, Petter, Hall, Lars, and Mercier, Hugo (2016). The Selective Laziness of Reasoning. *Cognitive Science*, 40(8), 2122-36.

Webber, Richard, Rallings, Colin, Borisyuk, Galina, and Thrasher, Michael (2014). Ballot Order Positional Effects in British Local Elections, 1973-2011. *Parliamentary Affairs*, 67, 119-36.

Wedgwood, Ralph (2017). The Value of Rationality. Oxford University Press.