episteme

ARTICLE

Social epistemology for individuals like us

Molly O'Rourke-Friel

Philosophy, Ursinus College, Collegeville, PA, USA
Email: mollyorourkefriel@gmail.com

Abstract
This paper argues that we are not just social epistemic creatures because we operate in social contexts. We are social epistemic creatures because of the nature of our epistemic cognitive capacities. In *The Enigma of Reason*, Hugo Mercier and Dan Sperber develop and defend the view that reasoning is a social competence that yields epistemic benefits for individuals through social interaction with others. I argue an epistemological consequence of their position is that, when beliefs are formed and sustained by dialogical deliberation, the relevant justification-conferring process doesn't occur solely within the cognition of the subject whose belief is under evaluation. Rather, it extends to include her interactive engagement with other deliberative participants. I argue this demonstrates that not all justification-conferring is evidential. As such, the analysis not only supports reconceiving the process reliabilist's notion of justification-conferring processes; it also serves as an argument against evidentialism. A goal of this paper is to demonstrate that social epistemology isn't merely a siloed offshoot of traditional epistemology. Even when approaching social epistemology using a conservative methodology, our investigation has serious implications for fundamental questions concerning epistemic normativity.

Keywords: Justification; reasoning; Mercier and Sperber; interactionist theory of reasoning; dialogue; reliabilism; evidentialism; social epistemology

1. Introduction

I sometimes joke that all social epistemology pieces start the same way: "Since Descartes, traditional epistemology has been individualistic with respect to [*insert epistemic phenomenon here*]. I will argue that [*epistemic phenomenon*] is in fact deeply social".[1] After this framing, the reader is usually confronted with one of two approaches to the notion of an epistemic phenomenon being "deeply social". Using Alexander Bird's terminology, we can refer to these two distinct approaches as "individual-social" and "social-social" (2010, 23).[2] The former approach attends to us as individual epistemic

---

[1]For a small sample of examples, see Goldberg 2010; Goldman 1999, 4; Audi 2005, 27; Lackey 2006, 1; Collin 2019, 21; Bird 2010.

[2]This distinction is similar to Goldman's distinction between *preservationist* and *expansionist* approaches to social epistemology (2010, 2).

subjects but gives special attention to our social epistemic contexts.[3] The latter treats social groups as epistemic subjects themselves.[4] Generally speaking, the following question determines the author's choice of approach: is the epistemic phenomenon *irreducibly* social?

Does this paper fall into that mold? Yes and no. This paper does fit the mold insofar as I am arguing for the following: "Since Descartes, traditional epistemology has been individualistic with respect to the relationship between reasoning and doxastic justification. I will argue that both are in fact deeply social . . . ". However, this paper doesn't quite fit the mold in another respect. This paper uses the individual-social approach given it concerns the doxastic justification of beliefs held by individuals. It also argues that doxastic justification is irreducibly social. This paper does not commit itself to the existence of group epistemic subjects or group knowledge in any important sense. Rather, it is intervening, or trying to dismantle an assumption of, the social-individual approach. We are not just social epistemic creatures because we operate in social contexts. We are social epistemic creatures because of the nature of our epistemic cognitive capacities.

Here is a sketch of how the paper will proceed: epistemologists largely follow Descartes' individualist tradition when it comes to our ability to reason. The standard view is that reasoning is a private competence, and the social enters into the epistemological picture when we start exchanging the outputs of that private competence with one another. However, new research puts pressure on this approach. In *The Enigma of Reason*, Hugo Mercier and Dan Sperber develop and defend an "interactionist account" according to which reasoning is a social competence that yields epistemic benefits for individuals through social interaction with others (2017, 9). I argue that the epistemological consequence of Mercier and Sperber's position is that, in cases where beliefs are formed and sustained by dialogical deliberation, the relevant process that confers justification on a deliberative participant's belief doesn't occur solely within the cognition of that particular subject. Rather, it extends to include her interactive engagement with other deliberative participants. A significant consequence of this claim is that not all that is justification-conferring is evidential. As such, the analysis not only supports reconceiving the process reliabilist's notion of justification-conferring processes; it also serves as an argument against evidentialism.

Before proceeding, I must address an important question: Why not take the social-social approach?[5] First, a goal of this paper is to demonstrate that social epistemology isn't merely a siloed offshoot of traditional epistemology. It doesn't leave the questions posed by traditional epistemologists untouched. It can't be accommodated merely by expanding our epistemic ontology to include, for example, social epistemic subjects/agents or group knowledge. Even when approaching social epistemology using a relatively conservative individual-social framework, our investigation has serious consequences for fundamental questions concerning epistemic normativity. I argue that if we posit a minimal desideratum on doxastic justification, one that ought to be acceptable to epistemologists working in the individualist tradition, we will see that

---

[3]E.g., Goldman 1999, Coady 1992. For an overview of this kind of approach, see also sections 3.1 and 3.2 of O'Connor, Goldberg, and Goldman 2024.

[4]E.g., see Bird 2010, 2014; Carter 2015; De Ridder 2014; Gilbert 1994, 2004; Klausen 2015; Hakli 2007; Lackey 2020; Pettit 2003; Schmitt 1994; Tuomela 1992, 2004.

[5]What follows distinguishes this project from social-social proposals that resonate with it, like Palermos' defense of collaborative knowledge using virtue reliabilism (2022). He argues that research on Transactive Memory Systems (TMS) and scientific collaboration shows that justification is irreducibly social. He then contends that this entails that there are group epistemic agents that possess collaborative knowledge.

appreciating the kind of social creatures we are demands adopting specific epistemic normative frameworks and theories (viz., externalism and reliabilism) and rejecting others (viz., internalism and evidentialism).[6]

Second, everything I argue for in this paper is consistent with the social-social approach writ large. For example, it could be the case that extended interactive belief-forming processes confer justification on beliefs held by individuals and, running parallel, there is social knowledge in Bird's functionalist sense or group belief in Margaret Gilbert's joint-commitment sense (1994, 2004). However, my argument does show that we can account for (at least part of) the inextricably social nature of our epistemic lives without having to adopt a theory of group belief, knowledge, or mind. My view also doesn't require that we posit different kinds of knowledge.[7] I contend this simplicity is a strength of the view. Do groups really have beliefs? Or are we speaking in metaphor and making generalizations when we talk as though they do? There are compelling arguments for thinking that metaphysics of group belief and knowledge is too burdensome to be plausible.[8] While some might argue that robustly social epistemology leaves all the intuitions and frameworks of individualist approaches behind, there is surely value in thinking about how we can accommodate what seems most plausible about the old individualist frameworks. "Individuals hold beliefs" is a likely candidate for inclusion in even the most thorough-going social epistemology.

2. The explanatory desideratum for theories of doxastic justification

I'll start by arguing that there is an *explanatory* desideratum that any satisfactory theory of doxastic justification must meet. This requirement becomes perspicuous if we disentangle ourselves from the weeds of specific theories and look at the bigger picture. What is a theory of doxastic justification trying to accomplish? When we are investigating doxastic justification, we aren't just thinking about the relationship between propositions and evidence. Rather, we are thinking about a particular subject's relationship to her beliefs. When we are investigating whether a subject is doxastically justified in their belief that *p*, we want to know that belief's *story*. How did the subject of the evaluation come to have this belief? Why does she still believe it?

Diving into the details of different theories, we find various positions on what precisely about a belief's story is epistemically relevant. Despite the meaningful differences between theories, there is something epistemologists of all stripes can sign onto: (non-)accidentality explanations are necessarily of epistemic interest. Consider, internalists and externalists, reliabilists and evidentialists alike can endorse the following: A true belief that is the product of wishful thinking isn't doxastically justified. True beliefs that are the result of wishful thinking are, in an important sense, only accidentally true. If a theory of doxastic justification fails to account for how wishful thinking diminishes a belief's justificatory status, we should think there is something deficient about that theory. Merely getting the correct "unjustified" verdict is insufficient; the epistemic unrespectability of wishful thinking, the idea that wishful thinking can only lead to true belief *on accident*, must function in the theory's explanation for why the subject's belief is unjustified.

These considerations point towards an explanatory desideratum on plausible theories of doxastic justification, one that stipulates that the correct theory will accurately account for all parts of the (non-)accidentality explanation for the truth or falsity of a

---

[6]Cf. Levy and Alfano (2020, 888) and Goldman (1999, 4).
[7]Cf. De Ridder and Palermos (2020).
[8]For examples, see Quinton 1976; Giere 2006, 2007; Thagard 1997, 2010.

subject's belief that *p*. In other words, all parts of the (non-)accidentality explanation for the truth or falsity of a subject's belief that *p* must accurately feature in the correct theory's account of the relevant belief's justificatory status.

A defense of this desideratum can be found in the distinction drawn between propositional and doxastic justification. When a subject has propositional justification, they are *in a position* to form a doxastically justified belief that *p*. In contrast, a subject is doxastically justified when their belief is properly based in some important respect. Contrast the case of wishful thinking discussed above with a case of expert consideration. Both the wishful thinker and the expert could be in possession of the same resources – access to the same evidence, discussions, tools, etc. In this case, both would be propositionally justified. Regardless, the wishful thinker does not have a doxastically justified belief because their belief isn't properly based on those epistemic resources.[9] The expert whose belief results from proper engagement with those resources and relevant discussions, who processes evidence well, uses the available tools reliably, etc. – this subject's belief is doxastically justified because of its epistemically appropriate history. When we look at the history of the expert's belief, we can see that it is not a mere accident that they arrived at the truth.[10]

This insight is used to undermine time-slice theories of justification. Time-slice theorists argue that we can determine whether a subject's belief is justified by taking a "snapshot" of the subject's cognition at the moment of belief evaluation.[11] But as our discussion of the wishful thinker shows, doxastic justification isn't just a matter of what epistemic resources one has at the moment of belief formation.[12] Doxastic justification is a matter of the *history* of belief formation: how the subject used those epistemic resources to produce or sustain the belief in question. Consideration of the belief's history is necessary for determining the (non-)accidentality of a belief's truth or falsity.[13]

One might argue that I need to say more about the nature of epistemically relevant accidentality. After all, in some ways of cashing out accidents, this desideratum is surely implausible. Consider a demanding view of epistemically relevant accidentality: a subject's true belief is a mere accident if anything other than the subject's epistemic agency is responsible for, or even partially caused, the subject to hold that belief. This view can't be right. Imagine an engaged student who is an astute reader of complex texts. She reads an article assigned by a professor and reliably forms a true belief about the article's thesis. We wouldn't want to say that her true belief is a mere accident and therefore unjustified because it was just good luck that she got off the course's waitlist.

---

[9]Importantly *proper* basing, not basing *simpliciter*, it a necessary condition for doxastic justification (Silva 2015). It is worth noting that the "standard" interpretation of the basing relation is causal (Korcz 2000, 526). See also Fantl (2019) for a discussion of the basing relation as a "diachronic requirement" (784-787).

[10]This point resonates with Sosa's discussion of accuracy, adroitness, and aptness (Sosa 2007, 22).

[11]Time-slice accounts of justification are often associated with internalism. For example, prominent defenders of time-slice views include Chisholm (1989, 59-60), Conee and Feldman (2004, 55, 101) – all defenders of internalism. However, see Fantl (2019) for discussions of how one could craft a time-slice externalist position (780-2) and for a defense of evidentialism as a historical account of justification (784-787). For an interesting discussion of different ways of conceptualizing historical justificational facts, see Kelly (2016, 44–51).

[12]This argument is made by Kornblith (1980, 602). While he does not construct his argument using the propositional vs. doxastic justification distinction, the introduction of these terms does not change the nature of the argument. Harman (1973, 30-33) makes a similar argument.

[13]This claim is not deeply controversial. The historical versus time-slice debate neither characterizes, nor tracks onto, the divide between internalists and externalists, or the divide between foundationalists and coherence theorists. Reliabilists and evidentialists alike can, with good reason, endorse the claim that justification has an important historical dimension.

Clearly, the influence of anything outside a subject's epistemic agency doesn't automatically generate a (non-)accidentality factor that must be included in a theory's justificatory status explanation.

For my purposes, a conservative account of epistemically relevant accidentality will suffice. Regardless of one's standing allegiance to a particular theory of doxastic justification, one can sign on to the following: features of a belief's story related to the (im)proper functioning of the subject's cognitive competences are relevant to the (non-)accidentality explanation. Even if we bracket the debate about whether influences external to the agent can generate (non-)accidentality factors that are relevant to doxastic justification, (non-)accidentality features embedded in the subject's own cognitive functioning are necessarily and always relevant. It is this minimal commitment that explains why the internalist, the externalist, the relaibilist and the evidentialist can all endorse the broad strokes explanation of the wishful thinker's lack of doxastic justification. We can amend the explanatory desideratum to account for this conservative commitment: all parts of the (non-)accidentality explanation for the truth or falsity of a subject's belief that p that are tied to her (im)proper cognitive functioning must accurately feature in the correct theory's account of the belief's justificatory status.

Before moving on, I want to identify an upshot of making this explanatory desideratum explicit: doing so orients our debates concerning theories of justification away from discussions about whether competing theories give the correct verdict in particular cases, to whether theories give complete and illuminating explanations of what is doing the justificatory work. With this in mind, I will transition to my discussion of Mercier and Sperber's interactionist theory of reasoning. In short, I argue that it is data about why subjects form non-accidentally true beliefs in social, deliberative settings and as such, it must be part of the justification story in relevant cases.[14]

3. Mercier and Sperber's interactionist theory of reasoning

In *The Enigma of Reason*, Mercier and Sperber (2017) develop and defend a novel "interactionist theory" of reasoning (9, 21, 182-183).[15],[16] Mercier and Sperber argue that proper analysis of the empirical research into the strengths and weaknesses of this capacity demonstrate that reasoning is primarily the exercise of a "social competence" (11). They are not claiming that reasoning fails to bring about "intellectual benefits" for individuals (11). Rather, their position is that reasoning brings about these individual epistemic and pragmatic benefits through "interactions with others" (11). Reasons, they argue, are produced for "social consumption": for convincing others and for justifying ourselves to others (127). When reasons are produced for this purpose, exchanged and evaluated in collaborative dialogic contexts, interlocutors more reliably reach true, well-reasoned conclusions. The epistemic gains of reasoning are explained by the interaction between interlocutors in dialogue, not interlocutors' individually reliable reasoning

---

[14]Going forward, unless otherwise stated, "justification" refers to "doxastic justification".

[15]See also Mercier and Sperber 2011 and 2012.

[16]By reasoning, Mercier and Sperber mean exercising the ability to produce, consciously entertain, and evaluate reasons for and against believing particular propositions and performing particular actions. On their view, a theory of reasoning requires an explanation of (i) what reasoning is for and, (ii) how it works. Mercier and Sperber's argue against the "intellectualist" theory of reasoning, a view that solely stakes out a position on (i): what reasoning is for. According to this traditional "dogma", reasoning is a capacity that helps us, *as individuals on our own*, form more true beliefs and act more pragmatically (Mercier and Sperber 2017, 9, 4, 330-331).

faculties. In other words, reasoning works well when collaboratively employed in the capacity's normal conditions: social and dialogical conditions (247).

Studies demonstrate that human reasoners systematically make certain kinds of errors. Reasoners commonly fail to abide by certain rules of probability and logic.[17] Studies also show that we are often mistaken about our own reasons for belief and action and frequently confabulate reasons (Wason and Evans 1975; Nisbett and Wilson 2002; Lucas and Ball 2005; Halberstadt and Wilson 2008; Hauser *et al* 2007; Carruthers 2011). Reasoners fall prey to the confirmation, or as Mercier and Sperber prefer "myside", bias: the tendency to single-mindedly pursue and consider only that evidence confirming the belief the reasoner already holds (Mercier and Sperber 2017, 213; Johnson-Laird and Byrne 2002; Stanovich and West 2008; Stanovich, West, and Toplak 2013). In addition to this bias in evidence and reason identification, individual isolated reasoners tend to be limited in their ability to produce compelling reasons for their own beliefs and to anticipate counterarguments (Mercier and Sperber 2017, 223; Kuhn 1991; Nisbett and Ross 1980; Perkins 1985). These last two traits of reasoning can lead to problematic "epistemic distortions": "overconfidence, polarization, belief perseverance" (Mercier and Sperber 2017, 246).

Roughly speaking, there are two schools of thought regarding the impact of these findings on our conception of human rationality. On one side of the debate are thinkers who take evidence of widespread systematic errors in human reasoning as a cause to be pessimistic about human rationality (e.g., Piattelli-Palmarini 1994, Tversky and Kahneman 1973); on the other, there are those who think there are alternative, more optimistic explanations (e.g., Pinker 1997, Gigerenzer 1991, 1998).[18] Mercier and Sperber, like many others sympathetic to the evolutionary psychologist's research program, reject a rationality-pessimistic explanation of the data. Their reasons for doing so are straightforward: such an explanation involves claiming that we have an ill-adapted cognitive capacity, which, from an evolutionary perspective, they think is hard to accept (Mercier and Sperber 2017, 331). If we accept the insights of evolutionary psychology, they argue we should think that an ill-adapted competence is puzzling and likely misdescribed. One does not start by stipulating that a certain system has a particular function, then look at evidence of that system failing to perform that function well as a cause for thinking that it is a poorly operating system. Rather, one looks at evidence of when and under what conditions the system works well, then one determines what function the system was adapted to perform. A well-adapted competence, they argue, is not an ability to perform a particular function under *any* conditions. Rather, a well-adapted competence functions well under *normal* conditions.

Mercier and Sperber argue that the data demonstrates reasoning works well under social, cooperative dialogical conditions, that is, when reasoners engage with, justify themselves to and disagree with other reasoners to coordinate beliefs and actions (Mercier and Sperber 2011, 247; 2017, 227, 183-186). Mercier and Sperber argue that the experiments used to justify rationality-pessimistic positions are not ones in which participants are in a "typical dialogical context" (Mercier and Sperber 2017, 227). In contrast, experiments that place participants in dialogical contexts demonstrate that

---

[17]For example, Tversky and Kahneman (1983) conducted research showing that human reasoners predictably fall prey to the conjunction fallacy (i.e., fallaciously believing that there is a greater probability of "A and B" than there is of just "A" or "B"). Moreover, research suggests that human reasoners fail at alarming rates to abide by logical rules governing conditionals (Mercier and Sperber 2017, 28; Byrne 1989; Evans 1989).

[18]See Mercier and Sperber 2017 for a placement of this discussion in the historical context of the mid-1970s "rationality wars" (21).

reasoners are quite adept at producing reasons to justify themselves *to their interlocutors* and crafting counterarguments to the positions held by their interlocutors (Mercier and Sperber 2017, 228; Resnick et al. 1993, 362-363; Kuhn, Shaw, and Felton 1997).

Mercier and Sperber argue that the characteristics some have taken to be evidence of reasoning's flaws – such as the ability to primarily produce reasons in defense of one's beliefs and the inability to anticipate counterarguments – are repurposed into collaborative strengths (2017, 247).[19] The confirmation or "myside bias" is less likely to result in "epistemic distortions" insofar as it functions as an effective division of cognitive labor (Mercier and Sperber 2017, 221). The bias divvies up the group's investigative work. Finding arguments and reasons in defense of a position with which one does not agree is cognitively costly. If interlocutors focus on producing arguments for their own position, the deliberative group can then pool their cognitive resources and increase their chances of arriving at a justified conclusion given that they are able to cover more evidential ground. Moreover, the biases that motivate the isolated individual are more likely to be corrected in dialogic contexts insofar as they are kept in check by the countervailing biases of interlocutors. Deployed in a collaborative dialogical context, the confirmation bias doesn't leave the individual reasoner only with arguments in favor of the position she came into the discussion believing. Rather, she has her own arguments, objections to those arguments leveled by her interlocutor, arguments for contrary positions and her own evaluation of those contrary arguments.

With this understanding of the conditions of proper functioning in place, the interactionist theory of reasoning takes shape: reasoning is not a capacity for individual epistemic gain through "solitary ratiocination" (Mercier and Sperber 2017, 218). When deployed collaboratively in social dialogic contexts, this *socially directed competence* brings about epistemic goods. The deployment of our cognitive capacity to produce justifying reasons and evaluate counterarguments is prompted, or tripped, when we are confronted with disagreement or doubt from others. It is the need to convince others— to convince them that what we say is true or that our actions are reasonable—that triggers our reasons-producing and evaluating faculties. Our individual reasons-producing and evaluating faculties, once tripped by others, are not by themselves reliable. Reliability results from combining our cognitive efforts. When I am forced to reevaluate the coherence of my reasons on the basis of your criticism, I may not do so reliably – I may be biased. When you consider the novel evidence I give you, you make not take it into consideration reliably – you might be biased. But through interaction, we bootstrap our way to epistemic goods.

4. A new kind of epistemic reliance

Let's say we accept the positions defended in the previous two sections: (i) the explanatory desideratum on theories of doxastic justification, and (ii) Mercier and Sperber's interactionist theory of reasoning. How ought epistemologists account for the epistemic gains that come from dialogical deliberation? Dialogical deliberation can surely be part of the history of a belief, but ought we conceive of it as a part of the belief's epistemically relevant history? Ought we think of dialogical deliberation as part of the relevant (non-)accidentality explanation for the truth or falsity of some beliefs? Can it be an element of a belief's history that impacts justification? In this section, I argue that we should answer all these questions in the affirmative.

---

[19]See also Paul Smart's discussion of "Mandevillian intelligence" (2018).

4.1. In defense of extended interactive belief-forming processes

Let's start by considering how the process reliabilist can handle the interactionist theory of reasoning. Process reliabilism, the view that justified beliefs are the outputs of reliable belief-forming processes, can elegantly account for epistemically relevant events that occur prior to the moment of belief formation. This is in large part because the theory was originally formulated to account for the intuition that a belief's justificatory status is impacted by that which "*causally initiates*" or "*sustains*" it (Goldman 1979, 8).[20] The general strategy of process reliabilism's inaugural text – Alvin Goldman's "What is Justified Belief?" – is to object to time-slice theories of justification by arguing that they yield unintuitive verdicts in cases where a belief is genetically corrupt (e.g., our case of the propositionally justified wishful thinker from section 2).

It is important to highlight that Goldman's process reliabilism, and the articulation of the view accepted by most of its proponents, maintains that the reliability of the *entire* process of belief formation is relevant to a belief's justificatory status. Flawless and reliable execution of modus ponens reasoning from the beliefs that p and *if p, then q* to the belief that q doesn't guarantee doxastic justification of the latter belief. Whether or not the belief that q is justified depends on whether the premise-beliefs are themselves reliably formed and sustained.[21] This insight prompts Goldman to reject what he terms "*Terminal-Phase Reliabilism*", the view that justification is conferred only by the reliability of the process that occurs at the very end of belief formation (16). Justification, according to the reliabilist, requires that the "*entire* history of the [belief-forming] process be sound (i.e., reliable or conditionally reliable)" (Goldman 1979, 16).

If Mercier and Sperber's interactionist theory is correct, process reliabilists should accept the following: Just as the processes by which a subject forms their premise-beliefs are relevant to the justificatory status of their conclusion-belief, interactive deliberation is relevant to the justificatory status of beliefs formed in dialogical contexts. When a subject engages in dialogical deliberation and forms a belief as a result of that engagement, part of what explains whether or not the belief was reliably formed is the subject's interaction with their interlocutors.[22] According to the interactionist theory, in these kinds of cases part of what accounts for the (un)reliability of subjects' reasoning is the (lack of) collaborative engagement with interlocutors. As such, any historical process reliabilist should accept that the dialogical deliberation is part of the epistemically relevant history of the belief and therefore is one of the processes, or part of the process, that confers justification. Once accepted, the process reliabilist must also grant that justification-conferring processes do not solely occur within individual subjects' cognitive systems. Rather, justification-conferring processes can extend beyond this boundary to include the cognitive systems of deliberative interlocutors. In cases of beliefs produced or sustained by dialogical deliberation,

---

[20]Goldman termed this early iteration of the view "*Historical Reliabilism*" (1979, 14). More precisely, the view is that a belief is justified if and only if it is the output of a reliable process, or a *conditionally* reliable process. A process is conditionally reliable when its inputs are beliefs. In such cases, the reliability of the process is conditional on the truth of the process inputs (1979, 13).

[21]This doesn't beg-the-question against the evidentialist or internalist, as it can be formulated in evidentialist terms. The evidentialist can say, "whether a conclusion-belief is justified is not just a function of whether the subject properly bases their belief on the evidence, where evidence is construed as the premises of their argument. The justificatory status of a conclusion-belief also depends on whether the premises of the argument are sufficiently evidentially supported."

[22]Like Goldman, I posit that justification is not merely conferred by processes that form beliefs but processes that sustain them. However, for brevity, I will continue using only the language of belief formation.

the epistemically relevant, justification-conferring process does not only occur in the cognition of the individual subject whose belief is under evaluation. Rather, the justification-conferring process extends beyond the individual's cognition, occurs in part in the cognitive systems of the deliberative dialogue's participants, and as such encompasses the back-and-forth of discussion. I will term this kind of process an *extended interactive belief-forming process.*[23]

4.2. An illustrative example

Let's take a case with sufficient engagement and adequate collaboration such that the deliberative dialogue is reliable, and (most of the) participants in the discussion come away with a true, justified belief. Imagine a psychology lab conducting a study on stigma. As they are interpreting their data, lab-member Tanja schedules a meeting to discuss a meta-analysis she believes must be taken into consideration. All the lab members look at the meta-analysis before coming to the meeting: some have beliefs about its relevance, while some don't yet have a settled view. In the back-and-forth of discussion, all are asked to defend the claims they make. The subsequent evaluation of reasons and arguments prompts the production of stronger arguments and the dismissal of weaker ones. For example, we can imagine that the following exchange is characteristic of the lab's deliberation.[24]

Kevin: Does that meta-analysis make a distinction between visible and concealable stigmatized identities?

Tanja: Given our focus is the effects of anticipated stigma, which is not a relevant distinction. We're interested in the effects of a person's beliefs about their own experience, not what other people can perceive.

Kevin: I disagree. The stressors of anticipated stigma with respect to concealable identities are different, as you must consider the consequences of when and whether to disclose your identity. That is relevant.

Tanja: And the person with a visible stigmatized identity has the stressor of not being able to conceal that identity. I think you are missing the point. I am not saying they aren't different stressors, but our research is about the effects of a person's *belief* that they will experience stigma.

Madeleine: I see what you are saying Tanja, but this meta-analysis also doesn't distinguish between studies from across a large timeframe. It obscures how stigma has changed in response to shifting legal policies. I don't think it bears much weight on our investigation.

---

[23]My proposal is compatible with de Ridder's notion of "scientific justification" (2014) insofar as we are both interested in capturing the collaborative nature of our epistemic achievements. Indeed, his view has a reliabilist component as well. However, de Ridder's theory isn't general; it specifically concerns the justification of "scientific knowledge", a "strong kind of reflective knowledge" that is explicitly internalist (45, footnote 15). My thesis makes no distinction between different types of knowledge and does not posit a group epistemic agent. This simplicity is a theoretical virtue of my proposal. Moreover, for the reasons discussed in section 5, I am concerned that internalist theories misunderstand what is fundamentally doing the justificatory work in cases of dialogical deliberation.

[24]Many thanks to Dr. Mora Reinka for her help in crafting this dialogue.

Tanja: If we don't take a nonsignificant correlation in the meta-analysis seriously, we risk of overlooking a moderator.

Kevin: But if the meta-analysis doesn't consider variables that are important to our study, then it isn't relevant. I agree with Madeleine.

The lab's deliberation continues in this manner. Most come to justifiably believe that the meta-analysis does not require them to reinterpret their study's results.

What is epistemically relevant to the justificatory status of this belief? What explains why the lab members' belief-forming processes are reliable? If we accept the interactionist theory, it is not merely a lab member's own isolated cognitive activity that determines the reliability of the process by which their belief was formed. Rather, their lab mates' cognitive activity and the interaction between that activity and their own individual processing, is also relevant. Tanja's critical evaluation of ideas raised by Kevin prompts Madeleine to consider other issues with the meta-analysis. This in turn prompts Tanja to produce further reasons for revising in light of the meta-analysis. The interactionist theory of reasoning tells us that these are not arguments, reasons and evaluations that the lab members would necessarily have produced working in isolation. Nor is it the case that the short processes yielding these intermediary arguments, reasons and evaluations are reliable when isolated from the interaction of the dialogue. Rather, these cognitive feats are tripped by one another's engagement and contributions. Moreover, these individual pieces of reasoning contribute to reliability when triggered by, and deployed in, this collaborative context. Therefore, the relevant justification-conferring process is social. For beliefs formed in this context, the processes extend beyond any one individual's cognition and include other interlocutors' cognitive activity as well as interactions between those cognitive events.

4.3. Problems for individualistic approaches

What happens if we try and craft a more individualistic epistemic accounting of the interactionist theory of reasoning? An individualistic accounting (a) isolates and evaluates solitary acts of reasoning that occur during the discussion and, (b) limits itself to the cognitive activity of the specific subject whose belief is being evaluated.

One might suggest that we use a process of aggregation: determining whether or not a lab member's belief is justified requires aggregating the reliability of their individual reasoning processes during discussion. This approach won't work, however. Remember, on the interactionist theory of reasoning, the individual, isolated process of ratiocination that produced intermediary arguments is likely not reliable when considered in isolation. Perhaps Tanja came to the discussion with an unjustified false belief about the importance of this meta-analysis, and this individual, isolated process of ratiocination is steered by confirmation bias. It is unreliable when evaluated on its own. Similar descriptions are true of other participants' segmented, individual processes. Why would aggregating individually unreliable processes yield a sufficiently reliable justification-conferring process?

More importantly, even if the reliability of individual pieces of ratiocination aggregated together to yield a sufficiently reliable process, this proposal should still be rejected. Such an account would characterize lab members who form justified beliefs as having done so *in spite of* their individual pieces of ratiocination that are unreliable in isolation. Tanja, for example, would be justified *in spite of* engaging in unreliable biased reasoning that produced intermediate arguments against the relevance of the meta-

analysis. This approach fails to satisfy the explanatory desideratum defended above. Why? Because on Mercier and Sperber's view, part of what explains Tanja and other lab members' success is that she engaged in this biased process of reasoning. Interlocutors that engage in group problem solving are epistemically successful *because of* these individual pieces of ratiocination that are unreliable in isolation. Although unreliable in isolation, they are, as discussed in section 3, integral for reliability in collaborative dialogue. As such, the individualistic analysis under discussion misrepresents epistemically relevant events in its accounting of the non-accidentality explanation.

One might propose an alternative individualistic explanation of the case: look at the precise moment when a lab member forms the belief that the meta-analysis doesn't bear on their study. Although the back-and-forth of dialogue is interesting, justification is conferred by the reliability of this end-of-discussion belief-forming process. This solution is implausible as it also fails to fully meet the explanatory desideratum. It doesn't account for cognitive activity on which the belief in question epistemically depends. According to the interactionist theory, the individual segments of solitary ratiocination (that are unreliable in isolation) are integral to the epistemic success of the collaborative dialogue, and as such, they are part of the epistemically relevant history of belief formation. In totally removing these cognitive events from the explanation, this individualist approach is embracing a kind of terminal-phase reliabilism that, as discussed above, violates the reliabilist's commitment to a historical theory of justification and the explanatory desideratum.

At this juncture, one might argue that there is at least one sense in which my own proposal, extended interactive justification-conferring processes, can be construed as individualistic. One might claim that the view requires looking at cognitive events that happen across individuals and "adding them up" to construct an extended interactive process. However, fundamentally my thesis concerns doxastic justification. My view is not that we should aggregate the *justificatory values* of cognitive events that happen across individuals to determine a belief's justificatory status. Moreover, any construal of my proposal as a mere aggregation of individual cognitive events is misleading. The relevant process isn't merely the sum of the cognitive events happening in individuals. It is also the interface between those events: the way they prompt and bump-up against one another. In my proposal, the interaction that occurs in between, at the intersection of the individual cognitive events, is part of the process.

This discussion makes clear the systematic issue individualistic approaches will have. Recall the minimal, conservative explanatory desideratum on justification defended above. This is the position that all parts of the (non-)accidentality explanation that are tied to the subject's (im)proper cognitive functioning must accurately feature in the belief's justificatory status. If we accept the interactionist theory of reasoning, we must accept that the epistemic non-accidentality generated by *proper* cognitive functioning can't be accounted for in an individualistic framework. The epistemic value of the conscious, reflective reasoning can't be understood by looking at the individual; understanding this kind of epistemically proper cognitive functioning requires looking at deliberative communities. A process reliabilist can account for this by positing that, in cases of dialogical deliberation, the relevant justification-conferring process extends beyond the individual to include the implicated cognitive activity of their interlocutors. In such cases, the relevant justification-conferring process is interpersonally, interactively extended.

4.4. Extended interactive justification-conferring processes clarified

The intricacies of the proposal are better understood by noting how it departs from previous versions of reliabilism. When first engaging with the question of how reliabilists should conceive of the "extent" of belief-forming processes, Goldman ("with some hesitation") suggests that we "restrict the extent of belief-forming processes to '*cognitive*' events, i.e., events within the organism's nervous system" (1979, 13). His primary reason for doing so is that justification is an evaluation of "how a cognizer deals with his environmental inputs" and we best capture this insight by restricting the relevant processes to that particular individual subject's cognitive system (1979, 13).[25] My proposal rejects this original reliabilist commitment. On my view, doxastic justification is not a pure metric of individual epistemic accomplishment: whether a subject's belief that *p* is justified isn't a function of how well her cognition alone dealt with her environmental inputs, because her cognition isn't "built" to deal with her environmental inputs alone. We must set aside a robust individual epistemic responsibility condition on justification. This is an upshot of accepting the interactionist theory: the proper functioning of human epistemic cognition cannot be understood by looking at an individual. The epistemic cognitive faculties we have don't merely operate in a social environment; they are themselves social.

My position echoes Sanford Goldberg's work in *Relying on Others*. In that text, Goldberg's central aim is to defend his "extendedness hypothesis": the view that the relevant justification-conferring process for beliefs formed on the basis of testimony includes not only the hearer's own cognitive processes, but "the cognitive processes implicated in the production of that testimony" (Goldberg 2010, 79). On Goldberg's view, whether the belief I form on the basis of testimony is justified depends on more than the reliability of my own belief-forming cognitive processes. It also depends on the reliability of the relevant cognitive processes occurring in the testifier. Goldberg's argument for this claim rests on the same commitments that prompt Goldman to adopt *historical* process reliabilism. Goldberg uses the notion of "epistemic reliance" to make this point (2010, 79). When we are evaluating justification, we care not only about what happens at the moment of belief formation but also the reliability of all of the processes on which the belief is epistemically reliant. Again, consider the case of wishful thinking above or instances of beliefs based on memories: If I base my belief that *p* on a memory of having seen that *p*, we care not only about the reliability of my memory recall but the reliability of the perceptual faculties that originally caused me to believe that *p*. Simply put, Goldberg's argument is that our dependence on other people in cases of testimonial exchange are similarly cases of epistemic reliance – not epistemic reliance on one's own temporally-prior cognitive processes, but epistemic reliance nonetheless. Where there is epistemic reliance, there is activity that is relevant to the justificatory status of a subject's belief (92). Insofar as cases of testimony involve subjects epistemically relying on one another, justification-conferring processes extend to account for that reliance, and as such include the implicated cognitive activity of testifiers.

Goldberg is arguing for his extendedness hypothesis on the grounds that it is necessary given the kind of epistemic reliance at play in cases of testimony. I am arguing

---

[25]It is important to note here that Goldman's focus is not merely on cognition happening within the minds of individuals. His focus is solely on cognition happening within the mind of the specific individual whose belief we are evaluating. Even in later work, Goldman is clear that he has an individualistic understanding of these processes: "Belief-forming and belief-revising processes are not themselves inherently social, and evidence used as a basis for forming and revising beliefs need not involve subject matter (i.e., content) concerning other people" (Goldman 2010, 6). My contention is that some belief-forming and justification-conferring processes are inherently social.

for the existence of extended interactive justification-conferring processes on the grounds that it is necessary given the kind of epistemic reliance at play in cases of dialogical deliberation.[26]

When a subject believes the conclusion of an argument, whether she is justified in believing the conclusion depends not only on the epistemic permissibility of her inference but her justification for believing each of the premises. She epistemically relies on historically prior epistemic events (that confer justification on her beliefs in the premises), and as such those prior epistemic events are part of what confers justification on her belief in the conclusion of her inference. The relevant justification-conferring process is *temporally* extended insofar as the subject is epistemically relying on historically prior processes. If we follow Goldberg, in straightforward cases of testimonial exchange, the recipient of the testimony epistemically relies on the testifier in forming a belief that *p* when the testifier asserts that *p*. The relevant justification-conferring process is *temporally* and *interpersonally* extended insofar as the subject is relying on a historically prior process that occurs in another epistemic subject. In dialogical deliberation, interlocutors epistemically rely on one another to trip reason-giving and reason-evaluating cognitive capacities. Interlocutors depend on one another to critically engage, to divide cognitive labor, so that the group can collectively navigate its way to the truth. When an interlocutor forms a belief as the result of social deliberation, the relevant justification conferring process is temporally, interpersonally and *interactively* extended.[27]

The kind of epistemic reliance we demonstrate in dialogical deliberation is meaningfully different from the epistemic reliance characteristic of straightforward cases of testimony, where this latter kind of case involves the preservation of content from one interlocutor to another.[28] In straightforward cases of testimony, we epistemically rely on

---

[26]In part, this section aims to show that if one accepts Goldberg's extendedness hypothesis, one ought to accept my proposal as well. However, I want to note that there are reasons one might accept my contention that justification-conferring processes extend interactively beyond individual cognizers in cases of dialogical deliberation but reject that the same is true in straightforward cases of testimony. In straightforward cases of testimony (in which the testifier asserts that *p* and the recipient of the testimony comes to believe that *p*), one could argue that the epistemic reliance at play can be completely justificatorily accounted for by evaluating the reliability of the testimony-recipient's judgment that their interlocutor is trustworthy. A similar explanation cannot be given in interactionist cases of dialogical deliberation. In such cases, it is not merely interlocutors' judgments of one another's reliability that yield epistemic gains, but rather the triggering or prompting effects interlocutors' contributions have on one another's faculties of reasoning.

[27]The extended interactive justification-conferring processes I defend here bear a resemblance to the "social belief-forming processes" Shieber proposes (2019). Our views converge insofar as we are both arguing that "the belief forming process that results in a believer's belief need not be limited to the cognitive processes of the believer alone" (93). However, there are important differences between our views, both in terms of how we are conceiving of these social processes, and our arguments in their defense. It is not clear that Shieber would endorse my proposal that processes can extend interactively to account for the reliability of deliberative dialogue (as opposed to merely the aggregate reliability of individual cognizers). Shieber argues (using both thought experiments and empirical data) that we should be suspicious that the justification of beliefs formed on the basis of testimony can be explained by our competence for assessing reliability. This, he argues, is particularly clear in cases where the testimony in question is the result of a "socially distributed systems of information transmission", largely because of the opacity of such systems as well as the "non-locality of expertise" (90). He is primarily concerned with social epistemic reliance on reliability, not the kind of interactive epistemic reliance I discuss here. Moreover, insofar as Shieber casts his discussion in terms of testimony, he is primarily concerned with accounting for the way we learn *from* others in those cases where the testifier is socially distributed. I am primarily concerned with accounting for the way we learn *with* others.

[28]A connection could be drawn to Goldberg's concept of an "epistemically engineered environment": "an environment that has been deliberately designed so as to decrease the cognitive burden on individual

the *reliability* of interlocutors' temporally prior cognitive processes. We rely on testifiers the same way we rely on thermometers; the fact that testifiers are cognizers, that this is a *social* interaction, plays no special or distinct role. In contrast, in dialogical deliberation, we rely on our interlocutors' participation to trigger in ourselves the kinds of cognitive activity that, when deployed in concert with others, allows a group to collaborate its way to a well-reasoned judgment. In such cases, an interlocutor epistemically relies on their deliberative peers to engage in a manner that trips reason-giving and reason-evaluating cognitive capacities. The individuals' reason-giving and reason-evaluating capabilities are not necessarily themselves reliable but contribute to reliability when deployed in dialogue with the joint cognitive activity of others.[29,30]

This is, of course, not to say that epistemic reliance on reliability is not a genuine and pervasive epistemic phenomenon. It is just to say that there is another, important kind of epistemic reliance; let's call it interactionist epistemic reliance.[31]

5. Against internalism and evidentialism

At this juncture, it is appropriate to question whether theories other than reliabilism can make similar accommodations. I argue that other prominent theories, namely internalism and specifically evidentialism, cannot. In essence, this section shows that making room for the social is a demanding task, one that bears on traditional questions about the nature of epistemic normativity.

Let's start by considering what resources are available to the access internalist. Access internalism is the view that justification is a function of whether a subject properly bases her belief on, or whether a subject's belief properly coheres with, other beliefs to which she has conscious, cognitive access (Bonjour 1980). An access internalist may try and

---

subjects in the attempts to acquire knowledge" (2020, 2795). Although it is certainly the case the Goldberg is investigating a kind of social epistemic reliance, his focus is not on the kind of epistemic reliance I have in mind. Goldberg is interested in a kind of reliance on others' epistemic good will, which is closely related to reliance on other's reliability – in particular, reliance on others to reliably set up environments to ensure or promote our epistemic success. This is distinct from *interactive* epistemic reliance.

[29]This view resonates deeply with Longino's account of objectivity in science, and the social nature of scientific knowledge (1990, 2002). Insofar as this argument turns on a psychological theory, rather than a sociological analysis of science, it provides new and distinct support for thinking about epistemic normativity as concerning social interactions, as opposed to discrete states of either individuals or groups.

[30]Levy and Alfano make a related point (2019). However, although they are arguing that we can gain "knowledge from vice", they are not defending a kind epistemic reliance distinct from dependence on interlocutor reliability (1). Rather, their contention is that that "some of our most significant epistemic achievements" result from behavior that looks epistemically vicious. Levy and Alfano's key examples are of intergenerational epistemic success, primarily cases in which humans are able to successfully navigate and flourish in challenging and diverse environments. They argue that these kinds of ecological epistemic successes result from an innate human disposition that is, at the individual level, an epistemic vice: extreme overimitation (7, 11). These cases are incredibly interesting and should prompt any reliabilist interested in social epistemology to think about how we ought to temporally carve up processes. Ultimately however, Levy and Alfano are arguing that behavior the virtue epistemologist would call vicious can be reliable, not that there is a kind of epistemic reliance distinct from dependence on interlocutor reliability.

[31]While my view can certainly find support in arguments for distributed cognition (e.g., 2007) and extended mind (e.g., Clark and Chalmers 1998), this paper does not require that one antecedently accept either hypothesis. My defense of extended interactive processes relies on the explanatory desideratum, which merely makes a claim that when evaluating justification, we must look at features of a belief's story that are related to the (im)proper functioning of a subject's epistemic cognitive competences. To accept this, one need not also accept that all that constitutes extended interactive justification-conferring process is cognition.

argue that they can account for the justificatory status of beliefs formed during dialogical deliberation by taking stock of the reasons and arguments deliberative participants consciously possess at the end of the deliberation. Returning to the psychology lab example sketched above, the access internalist might argue that at the end of their discussion, Madeleine, Tanja and Kevin not only have a belief that the meta-analysis isn't relevant, but they also possess a consciously accessible argument for that belief. After all, the back-and-forth of discussion has produced publicly articulated reasons and arguments, as well as evaluations of those reasons and arguments. The access internalist might say that it is interesting to acknowledge that these were generated in a social, dialogical fashion, but the justificatory status of the belief is conferred by the lab members basing this belief on whatever reasons they (consciously) possess at the end of the debate.

This analysis will not work. First, this view has the same issue as the individualist reliabilist strategies rejected above. By the lights of the interactionist theory, the pieces of reasoning that are considered "bad" when evaluated in isolation (e.g., biased reasoning) are part of the epistemically relevant explanation for why the beliefs formed in such settings are non-accidentally true. As such, these cognitive events are in some respect justification-conferring. Given the individual reasoning that is "bad" when considered in isolation plays no role in this access internalist accounting, the proposal falls short of accommodating the interactionist theory.

Second, Mercier and Sperber's view is not that conscious, reflective reasoning works well and evolved for use in conversation with others merely because it allows us to generate consciously accessible arguments in defense of the positions being discussed. No part of their view commits us to the idea that epistemic success involves the ability to recite arguments considered during the discussion after the conversation has concluded. This consideration also speaks against another possible access internalist proposal: that it is the ability to recite the whole deliberative encounter, the good and "bad" arguments alike, that does the justificatory work. Even if there are cases where interlocutors can recount some or all of the arguments discussed, we ought not think (as the access internalist would like us to) that it is the arguments' conscious accessibility that is conferring justification. The epistemic gains are the result of the dialogical interaction – the way in which dialogue *prompts* certain cognitive processes that aren't by themselves reliable but interact to form a process that is.[32] Like the individualist reliabilist strategies rejected above, the access internalist approach fails to satisfy the explanatory desideratum; it offers an incomplete and/or inaccurate explanation for why participants in sufficiently cooperative and collaborative deliberative dialogues non-accidentally form true beliefs.

---

[32]For further reasons to reject conscious-recitation proposals we can turn to Goldman's argument from "preservative memory" against time-slice theories of justification (Goldman 2009, 323). There are all sorts of facts we *know* even though we do not recall the evidential source that gave rise to the relevant belief, or the arguments we originally possessed in its defense. Despite not remembering the source, insofar as knowledge entails justification, we are justified in believing these propositions. This prompts Goldman to conclude that "[i]f S has a justified attitude D towards proposition P at t, and is S retains attitude D towards P until the later time t', via memory, then, *ceteris paribus*, S is still permitted to have attitude D towards P at t'" (Goldman 2009, 323). How does this relate to our discussion of the interactionist theory of reasoning? Even if we grant that Tanja, Kevin, and Madeleine formed consciously accessible arguments at the end of the discussion – indeed, even if they were able to recite the whole deliberative process, good and bad arguments alike – we should not think it is this ability for conscious recitation that is doing the justificatory work given it would be appropriate to say the following: at some time in the future, after they had forgotten those arguments, they are *still* justified in believing that Jesse is the most qualified candidate.

What about mentalist internalists? Do they fare better? The mentalist internalist argues that the justificatory status of a subject's belief supervenes on that subject's mental states. Mentalist internalists are traditionally time-slice theorists. To figure out the justificatory status of a subject's belief, we look that the subject's mental states *at the time of belief evaluation*. Consider Conee and Feldman's characterization of evidentialism, a species of mentalist internalism: "Doxastic attitude *D* toward proposition p is epistemically justified for S at *t* if and only if having *D* toward *p* fits the evidence S has at *t*" (Conee and Feldman 1985, 15). Given the theory posits that a belief's justificatory status can be determined by taking a "snapshot" of the subject's mental states at the moment of belief evaluation, this position is incompatible with accepting that justification has an integral historical dimension. As such, like alternative proposals considered above, traditional evidentialism will leave the pieces of reasoning that are considered "bad" when evaluated in isolation out of the justificatory story in cases where beliefs are the result of deliberative dialogue. Evidentialism fails to satisfy the explanatory desideratum, too.

That said, internalist evidentialists can resist the time-slice characterization of their view and argue that the fundamental insights of the theory can be preserved while accepting a historical dimension to justification. For example, one could argue that evidentialism requires that subjects properly base their beliefs on their evidence, and there is a "diachronic requirement" on proper basing (Fantl 2019, 784-786). The central idea is that proper basing takes time: for your belief that *p* to be properly based on your evidence *e*, *e* has to cause you to believe (or to sustain the belief) that *p*. So long as the evidentialist has an argument for the claim that epistemic causes always precede their effects, they can reject time-slice epistemology. Does *historical* evidentialism, as a species of mentalist internalism, have the resources to accommodate the epistemic consequences of Mercier and Sperber's interactionist account?

It does not. Seeing this requires taking a step back and looking more broadly at the criticisms I have been making of individualist and internalist approaches to handling the epistemic consequences of Mercier and Sperber's interactionist theory. Again: evaluations of justification are ultimately a matter of figuring out what about the subject's proper cognitive functioning makes it the case that their belief is non-accidentally true. The interactionist theory of reasoning has the following epistemic consequence: Part of what explains a subject's proper cognitive functioning in (epistemically healthy) social, deliberative contexts are segments of individual participants' reasoning that are "bad" when considered on their own, separate from the rest of the interaction. Therefore, (at least part) of what justification is conferring in these cases is not evidential.

The historical evidentialist might try and insist that the role of this "bad" reasoning is in some sense accounted for: after all, the "bad" reasoning prompts the production of good evidence, and evidence is at the heart of the justificatory story on their view. However, this response will not do. Consider two groups, one that engages in collaborative, cooperative reliable dialogue, another that engages in unreliable dialogue (perhaps as a result of there not being sufficient disagreement between group members, which can lead to groupthink). By chance, the unreliable group happens to get lucky on this occasion. Despite not engaging collaboratively and cooperatively with one another and interacting in a manner that generally leads to poor epistemic outcomes, at the end of their discussion the deliberative participants base their true beliefs on the same evidence as the group that reasoned reliably. Given the two groups end up with the same evidence, the historical evidentialist can't account for the intuition that members of the collaborative group that engaged in disagreement, but not the groupthink contingent, have justified beliefs at the end of their deliberation.

One might argue that the evidentialist can push back by appealing to the idea that justification requires the absence of defeaters. The evidentialist could claim that the groupthink contingent has a defeater: the lack of collaborative dialogue. I do not think this strategy amounts to a persuasive defense. What explains the epistemic success of the collaborative group whose members gain justified beliefs? It isn't higher-order evidence about the reliability of their discussion, and it isn't evidence that there was dialogic collaboration. Rather, it is the interactive dialogue itself; the reasons-prompting and evaluating function that interlocutors play. To the evidentialist pushing this objection, I would return to the explanatory desideratum defended above. Cashing out the justificatory role of deliberative dialogue merely in terms of evidence offers an inapt (non-)accidentality explanation of these kinds of cases.[33]

This discussion demonstrates the limits of internalist and evidentialist theories to account for the interactionist theory of reasoning, and indeed, the epistemic role of the social. These theories construe "the social" as an environment that impacts individuals, and the epistemic role of the social is accounted for by evaluating those individual impacts. A plausible theory of doxastic justification must acknowledge that some epistemically relevant proper human cognitive functioning is itself inextricably social – it isn't merely a matter of a social context.

6. Anticipating objections

In this section, I anticipate and respond to two potential objections. The first contends that deliberative dialogue is better thought of as a context in which a process is used rather than as a constituent part of a justification-conferring process. The second charges my argument with conflating the distinction between what causes a subject to form a belief and what confers justification on a belief.

6.1. Contexts versus processes

Many reliabilists will plausibly argue that when evaluating justification, we look at the reliability of a belief-forming process relative to a particular context. On this view, whether a belief is justified depends on whether the relevant token belief-forming process is "reliable in environments of the same types as the one in which the belief was formed" (Heller 1995, 504).[34] In line with this commitment, one might argue that dialogical deliberation is not a constituent part of a justification-conferring process, rather it is a context in which a justification-conferring process can take place – namely, a context in which conscious, reflective reasoning is reliable according to the

---

[33]This alternative proposal also arguably rests on an implausible account of normative undercutting defeaters (i.e., higher-order evidence of one's (un)reliability that one ought to have considered; Lackey 2006). Evidence of non-deliberation arguably cannot be persuasively understood as this kind of defeater. If evidentialism committed itself to the view that justification requires defeating all possible undercutting defeaters, it would find itself facing a regress problem and on the road to skepticism. We need to say something about when an undercutting defeater is sufficiently relevant. Ought the group have known that the lack of diverse views rendered their dialogue unreliable? Given the historical dominance of the intellectualist theory of conscious reflective reasoning (see footnotes 16 and 18), this seems implausible. Indeed, against the backdrop of contemporary everyday calls to "do our own research", I struggle to have the intuition that the group ought to have considered that deliberation contributes to reliability.

[34]Heller defends this idea in responding to the generality problem for, arguing that it is not a substantive concern so long we take a contextualist approach to evaluations of reliability. Goldman (1979, 86-91) argues that the reliability of process-types is judged relative to particular contexts of use in crafting a relevant-alternatives response to the well-known barn facades case.

interactionist theory. One might argue that to include dialogical deliberation as part of a justification-conferring process would be akin to mistakenly including the quality of light, or the functioning of a light source, as part of the process that confers justification on visual perceptual beliefs. Yes, the functioning of a light source is part of the history of belief formation. However, the proper way to account for its epistemic relevance is to recognize that (human) visual perceptual processes are not reliable *simpliciter*, but rather reliable in contexts with sufficient/adequate light. One might object to my argument by claiming that, similarly, the reliability of certain cognitive capacities is determined relative to the social features of one's context. To oversimplify, the idea would be roughly: conscious, reflective reasoning capacities are reliable in contexts with sufficient amounts of deliberative, interactive dialogue and unreliable in contexts in which individuals are isolated.

This prima facie plausible position maintains the individualist intuition that justification-conferring processes are confined to an individual's cognitive system. Why argue for the more extreme conclusion that there are extended interactive justification-conferring processes if it is not necessary? We need to accept the more extreme conclusion because this dialogue-as-context position doesn't sufficiently capture the phenomenon as Mercier and Sperber have described it. Recall the discussion from section 3.2. Mercier and Sperber's claim is not that individuals' reliable solitary reflective capacities are tripped when engaging in dialogical deliberation. Rather, their claim is that our capacity to defend our beliefs, to produce reasons in support of our beliefs and to evaluate interlocutors' arguments is tripped in dialogic deliberation. These competences are not individually truth-tracking. Rather, deployed collectively, they help deliberating groups reach the truth. Insofar as we don't become individually more reliable in dialogical contexts, we can't accommodate the interactionist theory by positing that deliberative dialogue is a context relative to which individual conscious reflective reasoning must be evaluated.

6.2. Causation versus justification

Some might object to my defense of extended interactive belief-forming processes by asserting that even a proponent of historical accounts of justification is not committed to saying that *everything* in a belief's causal history is epistemically relevant – that is, justification-conferring or justification-detracting. For example, one could argue that both my drinking a cup of coffee this morning and the processing of my visual perceptual faculties are causally related to my present justified belief that my cat is asleep in the window. My cognitive processes would undoubtedly have functioned differently had I not consumed coffee this morning, likely more sluggish, and therefore my drinking coffee this morning had some causal impact on my belief about the cat's whereabouts. Despite this causal connection, most proponents of historical accounts of justification think it's inappropriate to say that coffee consumption is part of the process that confers justification on my belief. We must distinguish between that which *causes* a belief and that which *confers justification* on a belief.

Just as my coffee consumption is irrelevant to the justificatory status of my belief concerning my cat's whereabouts, one might argue that deliberative dialogue is similarly epistemically irrelevant to beliefs I form in discussion with others. Deliberative dialogue may be part of a belief's causal history, but that doesn't mean it is necessarily epistemically relevant. If we ought not think of deliberation and interaction as epistemically evaluable parts of the causal histories of beliefs, then, of course, we should abandon the idea that they are part and parcel of processes that confer justification.

In responding to this objection, it is important to think about what explains the intuition that drinking coffee is not part of an epistemically evaluable process despite the effect it has on an epistemically evaluable process. Why is the operation of our visual perceptual system epistemically evaluable or justification conferring, but not the process

of drinking coffee? Here is one explanation: It is appropriate to epistemically evaluate visual perceptual processes, and understand them as justification-conferring, because our visual perceptual systems evolved to help us form accurate beliefs about the world around us. Accurate belief formation is what our visual cognitive processes are for; the teleological role of our visual perceptual faculty is distinctly epistemic. In contrast, the way in which coffee is metabolized does not have the same distinctly epistemic teleological role of improving the reliability of cognition. Drinking and metabolizing coffee, like metabolizing other sources of sustenance, is a process of energy conversion that supports general metabolic function. It serves many functions in addition to epistemic cognitive functioning. Insofar as coffee consumption is lacking a distinctly teleological-epistemic watermark, it cannot be said to be part of, or a distinct, justification-conferring process.

Given this analysis, what should we say about the social dialogical interactions that are the focus of this paper? Do they have the required teleological-epistemic watermark necessary for a process to be evaluated as justification conferring? The answer is yes. As is discussed above, Mercier and Sperber's defense of their interactionist theory of reasoning is a teleological analysis. Their argument proceeds by looking at the empirical data about our ability to reason and then formulating an explanation of that data that makes evolutionary sense insofar as it characterizes reasoning as a *successful* epistemic function. On examining the data, what they determine is that the best way to understand conscious, reflective reasoning as properly adaptive cognitive functioning is as part of our social epistemic lives. Conscious reflective reasoning is instrumental to epistemic success when deployed in interactive engagement with others. The claim is that conscious, reflective reasoning is adapted for use in deliberative interaction because that is when it leads to epistemic success.

Mercier and Sperber position their interactionist theory of reasoning as a replacement for the classical intellectualist theory of reasoning. The intellectualist theory posits that conscious reflective reasoning is a cognitive function with an epistemic teleological watermark, but one that is designed for individuals to deploy in isolation. Mercier and Sperber's central criticism of this view is that, given the epistemic errors that conscious reflective reasoning exemplifies when deployed in isolation, it doesn't make sense to identify isolated, individual epistemic gain as the adaptive function of this kind of cognition. The central advantage of the interactionist theory of reasoning is that it gives an explanation of conscious reflective reasoning that is compatible with the empirical data on when this kind of reasoning reliably leads to epistemic success. The result is that conscious reflective reasoning in social dialogical interaction can be said to be an adaptive ability with a distinctly epistemic function in the same sense that our visual perceptual processes have a distinctly epistemic function. As such, it would be misguided to classify interactive dialogue and deliberation with non-epistemic causal influences on our belief forming processes, like coffee-drinking.

One other reply to this objection is necessary. Coffee-drinking is part of the causal history of belief-formation insofar as it has an impact on the reliability of an *individual* cognizer's cognition. When we are insufficiently caffeinated or sleep-deprived, we become, as individuals, less reliable. To the extent that anyone would want to account epistemically for coffee-drinking, it seems that it would be appropriate for them to do so by discussing the contexts or environments in which the processes occur. Sleep deprivation may result in blurry vision, but we don't want to say that, as a result, visual perpetual processes are unreliable and can't yield justified beliefs. Rather, we should say that visual perceptual processes are reliable relative to contexts with sufficient lighting and when carried about by a sufficiently alert human subject. But recall from the discussion in section 4.2, insofar as

dialogical deliberation doesn't make individual reasoning more reliable, we can't accommodate the interactionist theory by similarly discussing proper contexts of use.

7. Conclusion

I have argued that if Mercier and Sperber's interactionist theory is true, process reliabilists should accept extended interactive justification-conferring processes. Moreover, this amounts to an argument for process reliabilism, as internalism in general and evidentialism in particular, are unable to account for interactionist reasoning. Social epistemology requires normative frameworks that can account for the way in which our epistemic cognition can be inextricably social.

To be clear, my arguments do not commit us to thinking that solitary, non-deliberative conscious reflective reasoning never plays a positive justification-conferring role. Consider, we can accept that our visual faculties are adapted to function optimally in contexts with sufficient lighting without also thinking that any visual belief formed when those faculties operate less reliably (e.g., in dimly lit settings) is unjustified. Moreover, we can think that isolated ratiocination can play a positive justificatory role while thinking that, in light of the arguments above, traditional epistemology has historically mischaracterized its role in our epistemic lives.[35] This clarification is emblematic of the paper's upshot. Social epistemology is not an epistemological sub-field. Rather, it is a methodological commitment that should reorient our attitudes towards the questions posed, and the answers offered by, traditional, individualistic epistemology.

Consider how, beyond accommodating Mercier and Sperber's interactionist theory of reasoning, extended interactive justification-conferring processes can be put to other important work. They can offer interesting explanations of social epistemic phenomena like echo chambers. Insofar as interactionist reasoning thrives on prompting mechanisms that require disagreement, we should anticipate epistemic distortions when viewpoints are "actively excluded" and "discredited" (Nguyen 2020, 1). With interactively extended justification-conferring processes, we can explain why echo-chamber-captives are saddled with false beliefs and vindicate the intuition that they are unjustified. More importantly, interactively extended justification-conferring processes draw our ameliorative focus to the appropriate place: the health of our epistemic communities. The health of our epistemic communities isn't a function of our individual reliability but rather the interactive deliberative dynamics. It suggests that what we want isn't for our fellow epistemic community members to be perfectly responsive to their evidence but rather up for the task of engaging with one another.

This is not a false dichotomy. Efficient division of cognitive labor that contributes to the reliability of dialogue involves subjects exercising biases (see sections 3 and 4). As such, we are facing a genuine trade-off. Perhaps one might argue that if epistemic subjects were perfectly responsive to their evidence, then there wouldn't be a need for a division of cognitive labor that requires subjects to be less than perfectly responsive to evidence. Alternatively, one could argue that we could get the gains of deliberative dialogue by having epistemic subjects assume the role of biased reasoners in dialogic contexts, all while genuinely withholding judgment until they can form a judgment that perfectly responds to the evidence. However, both strategies would be resisting epistemic

---

[35]See Kornblith (forthcoming) for a similar argument.

normative theorizing that is responsive to the epistemic creatures we are, not the idealized ones we might wish to be. Surely, this commitment is part of the ethos of social epistemology: epistemic normative theorizing should attend to the kinds of epistemic creatures we are, and we are social epistemic creatures.[36]

References

Audi R. (2005). 'The epistemic authority of testimony and the ethics of belief'. In A. Dole and A. Chignell (Eds.), *God and the Ethics of Belief: New Essays in Philosophy of Religion*. Cambridge: Cambridge University Press.

Bird A. (2010). 'Social Knowing: The Social Sense of 'Scientific Knowledge'.' *Philosophical Perspectives* **24**(1), 23–56.

Bird A. (2014). 'When is There a Group that Knows? Distributed Cognition, Scientific Knowledge, and the Social Epistemic Subject.' In J. Lackey (Ed.), *Essays in Collective Epistemology* (42–63). Oxford: Oxford University Press.

Bonjour L. (1980) 'Externalist Theories of Empirical Knowledge.' *Midwest Studies in Philosophy* **5**, 53–73.

Byrne R.M. (1989). 'Suppressing Valid Inferences with Conditionals.' *Cognition* **31**(1), 61–83.

Carruthers P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Carter, J.A. (2015). Group Knowledge and Epistemic Defeat. *Ergo: An Open Access Journal of Philosophy* **2**(28), 711–735.

Chisholm R. (1989). *Theory of Knowledge*. Englewood Cliffs, NJ: Prentice Hall.

Clark A. and Chalmers D. (1998). 'The Extended Mind.' *Analysis* **58**(1), 7–19.

Coady C.A.J. (1992). *Testimony: A Philosophical Study*. Oxford University Press.

Collin F. (2019). 'The Twin Roots and Branches of Social Epistemology.' In M. Fricker, P. Graham, D. Henderson and NJ Pedersen (Eds.), *The Routledge Handbook of Social Epistemology* (21–28). Routeledge.

Conee E. and Feldman R. (2004). *Evidentialism: Essays in Epistemology*. New York: Oxford University Press.

De Ridder J. (2014). 'Epistemic Dependence and Collective Scientific Knowledge.' *Synthese* **191**, 37–53.

Evans J.B.T. (1989). *Bias in Human Reasoning: Causes and Consequences*. Hillsdale: Larence Erlbaum Associates, Publishers.

Fantl J. (2019). 'Evidentialism as an Historical Theory.' *Australasian Journal of Philosophy* **98**(4), 778–791.

Giere R. (2006). *Scientific Perspectivism*. Chicago: University of Chicago Press.

Giere R. (2007). Distributed cognition without distributed knowing. *Social Epistemology* **21**(3), 313–320.

Gigerenzer G. (1998). 'Ecological Intelligence: An Adaptation for Frequencies.' In D.D. Cummins and C. Allen (Eds.), *The Evolution of Mind*. New York: Oxford University Press.

Gigerenzer G. (1991). How to Make Cognitive Illusions Disappear: Beyond 'Heuristics and Biases.' *European Review of Social Psychology* **2**(1), 83–115.

Gilbert M. (2004). 'Collective Epistemology.' *Episteme* **1**(2), 95–107.

Gilbert M. (1994). Remarks on Collective Belief. In F.F. Schmitt (Ed.), *Socializing Epistemology: The Social Dimensions of Knowledge*, 235–256. Landham, MD: Rowman & Littlefield.

Goldberg S.C. (2010). *Relying on Others: An Essay in Epistemology*. Oxford: Oxford University Press.

Goldberg S.C. (2020). 'Epistemically Engineered Environments.' *Synthese*, 2983–2802.

Goldman A. (1979). 'What is Justified Belief?' In George S. Pappas (Ed.), *Justification and Knowledge: New Studies in Epistemology*. Boston, MA: D. Reidel Publishing Company.

Goldman A. (1999). *Knowledge in a Social World*. Oxford: Oxford University Press.

Goldman A. (2009). 'Internalism, Externalism, and the Architecture of Justification.' *The Journal of Philosophy* **106**(6), 309–338.

Goldman A. (2010). 'Why Social Epistemology is Real Epistemology.' In A. Haddock, A. Millar and D. Pritchard (Eds.), *Social Epistemology*. Oxford: Oxford University Press.

Hakli R. (2007). On the possibility of group knowledge without belief. *Social Epistemology* **21**(3), 249–266.

Halberstadt J. and Wilson T. (2008). 'Reflections on Conscious Reflection: Mechanisms of Impairment by Reasons Analysis.' In J.E. Adler and L.J. Rips (Eds.), *Reasoning: Studies of Human Inference and its Foundations* (548–565). New York: Cambridge University Press.

Harman G. (1973). *Thought*. Princeton, NJ: Princeton University Press.

Hauser M., Cushman F., Young L., Kang-Xing Jin R. and Mickhail J. (2007). 'A Dissociation between Moral Judgements and Justifications.' *Mind and Language* **22**, 1–21.

Heller M. (1995). 'The Simple Solution to the Problem of Generality.' *Noûs* **29**(4), 501–515.

Johnson-Laird P.N. and Byrne R.M.J. (2002). 'Conditionals: A Theory of Meaning, Pragmatics, and Inference.' *Psychological Review* **109**, 646–678.

Kelly T. (2016). 'Historical versus Current Time Slice Theories in Epistemology.' In B.P. McLaughlin and K.Hilary (Eds.), *Goldman and His Critics* (43–65). Malden, MA: Wiley Blackwell.

Klausen S.H. (2015). 'Group knowledge: A real-world approach.' *Synthese* **192** (3), 813–839.

Korcz KA. (2000). 'The Causal-Doxastic Theory of the Basing Relation.' *Canadian Journal of Philosophy* **30**(4), 525–550.

Kornblith H. (1980). 'Beyond Foundationalism and Coherence Theory.' *Journal of Philosophy* **77**(10), 597–612.

Kornblith H. (forthcoming). The Social Practice of Giving and Asking for Reasons.' In Preston Stovall and Ladislav Koreň (Eds.), *Why and How We Give and Ask for Reasons*. Oxford: Oxford University Press.

Kuhn D. (1991). *The Skill of Arguments*. Cambridge: Cambridge University Press.

Kuhn D., Shaw V., and Felton M. (1997). 'Effects of Dyadic Interaction on Argumentative Reasoning.' *Cognition and Instruction* **15**(3), 287–315.

Lackey J. (2006). 'Knowing from Testimony.' *Philosophy Compass* **1**(5), 432–438.

Lackey J. (2020). *The Epistemology of Groups*. New York: Oxford University Press.

Levy N. and Alfano M. (2019). 'Knowledge from Vice: Deeply Social Epistemology.' *Mind* **129**(515), 887–915.

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

Longino H. (2002). *The Fate of Knowledge*. Princeton, NJ: Princeton University Press.

Lucas E.J. and Ball L.J. (2005). 'Think-Aloud Protocols and the Selection-Task: Evidence for Relevance Effects and Rationalizations.' *Thinking and Reasoning* **11**, 35–66.

Mercier H. and Sperber D. (2011). 'Why do Humans Reason? Arguments for an Argumentative Theory.' *Brain and Behavioral Sciences* **34**, 57–111.

Mercier H. and Sperber D. (2012). 'Reasoning as a Social Competence.' In Hélène Landemore and Jon Elster (Eds.), *Collective Wisdom: Principles and Mechanisms* (368–392). Cambridge: Cambridge University Press.

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

Nguyen T. (2020). Echo Chambers and Epistemic Bubbles. *Episteme* **17**(2), 141–161.

Nisbett R. and Ross L. (1980). *Human Inference: Strategies and shortcomings of social judgement*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett R. and Wilson T. (2002). 'Telling More than We Can Know: Verbal Reports on Mental Processes.' *Psychological Review* **85**, 231–259.

O'Connor C., Goldberg S. and Goldman A. (2024). 'Social Epistemology'. In E.N. Zalta and U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy (Summer 2024 Edition)*. https://plato.stanford.edu/archives/sum2024/entries/epistemology-social/

Palermos S.O. (2020). 'Epistemic Collaborations: Distributed Cognition and Virtue Reliabilism.' *Erkenntnis* **87**, 1481–1500.

Perkins D.N. (1985). 'Postprimary Education has Little Impact on Informal Reasoning.' *Journal of Educational Psychology* **77**, 562–571.

Pettit P. (2003). Groups with Minds of Their Own. In F.F. Schmitt (Ed.), *Socializing Metaphysics: The Nature of Social Reality*, 167–193. Lanham, MD: Rowman & Littlefield Publishers.

Piattelli-Palmarini M. (1994). *Inevitable Illusions: How Mistakes of Reason Rule Our Mind*. New York: Wiley.

Pinker S. (1997). *How the Mind Works*. New York: W. W. Norton and Company.

Quinton A. (1976). 'Social Objects.' *Proceedings of the Aristotelian Society* **75**(1), 1–27.

Resnick L.B., Salmon M., Zeitz C.M., Wathen S.H. and Holowchk M. (1993). 'Reasoning in Conversation.' *Cognition and Instruction* **11**(3–4), 347–364.

Schmitt F. (Ed.) (1994). *Socializing Epistemology: The Social Dimensions of Knowledge*. Lanham, MD: Rowman & Littlefield.

Shieber J. (2019). 'Socially Distributed Cognition and the Epistemology of Testimony.' In M. Fricker, P. Graham, D. Henderson and N. Jang Pedersen (Eds.), *The Routeledge Handbook of Social Epistemology* (87–95). New York: Routledge.

Silva P. (2015). 'On Doxastic Justification and Properly Basing One's Beliefs.' *Erkenntnis* **80**(5), 945–955.

Smart P.R. (2018). 'Mandevillian Intelligence.' *Synthese*, 4169–4200.

Sosa E. (2007). *A Virtue Epistemology*. New York: Oxford University Press.

Stanovich K.E. and West R.F. (2008). 'On the Failure of Cognitive Ability to Predict Myside and One-sided Thinking Biases.' *Thinking and Reasoning* **14**(2), 129–167.

Stanovich K.E., West R.F. and Toplak M.E. (2013). 'Myside Bias, Rational Thinking, and Intelligence.' *Current Directions in Psychological Science* **22**(4), 259–264.

Thagard P. 1997). Collaborative Knowledge. *Noûs* **31**(2), 242–261.

Tuomela R. (1992). 'Group Beliefs.' *Synthese* **91**(3), 285–318.

Tuomela R. (2004). 'Group knowledge analyzed.' *Episteme* **1**(2), 109–127.

Tversky A. and Kahneman D. (1973). 'Availability: A Heuristic for Judging Frequency and Probability.' *Cognitive Psychology* **5**(2), 207–232.

Tversky A. and Kahneman D. (1983). 'Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement.' *Psychological Review* **90**(4), 293–232.

Wason P. and J.B.T. Evans (1975). 'Dual Processing in Reasoning.' *Cognition* **3**, 141–154.

Molly O'Rourke-Friel is an assistant professor of philosophy at Ursinus College. From 2023 to 2024 she was a visiting fellow at the Applied Epistemology Project at The University of North Carolina Chapel Hill. She received her PhD from the University of Massachusetts Amherst in the summer of 2022. She works on issues in social and applied epistemology, as well as philosophy of mind.