Belief and Ameliorative Epistemology

Citation

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:33840684

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Belief and Ameliorative Epistemology

A dissertation presented
by
Emily C. McWilliams
to
The Department of Philosophy
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
in the subject of
Philosophy

Harvard University
Cambridge, Massachusetts
August 2016
My dissertation is in three parts. In “Evidentialism and Belief Polarization,” I consider the epistemic import of a belief revision process known as belief polarization, in which exposure to a mixed batch of evidence reliably causes people to increase confidence in whatever their antecedent belief was. I argue against Tom Kelly's claim that the beliefs that result from this process are justified on evidentialist grounds, and I take stock of what the phenomenon can teach us about evidentialism as a theory of justification. In “Access Is Not Necessary for Justification,” I focus on a debate between mentalism and accessibilism, as internalist theories of epistemic justification. I argue against the received view that mentalism, as distinct from accessibilism, is not worth defending. In “Epistemic Oppression and Ameliorative Epistemology,” I offer an account of a phenomenon that I call epistemic oppression. It occurs when members of a social group are subject to a systematic and unfair disadvantage within a social epistemic structure, resulting from the operation of a capacity to control exercises of their epistemic agency, thus systematically and unfairly limiting their epistemic flourishing. I argue that the concept of epistemic oppression helps us to identify and understand the kind of wrong that happens in these cases where an agent's capacity for epistemic flourishing is unfairly impeded.
# Table of Contents

**Introduction** .................................................................................................................. 1

**I. Evidentialism and Belief Polarization** ........................................................................... 6
1. Introduction ....................................................................................................................... 6
2. Kelly on The Psychological Phenomena That Underwrite Polarization ......................... 7
3. Kelly’s Normative Assessment .......................................................................................... 8
4. Normative (Re)assessment ............................................................................................... 10
   4.1 Locating the Problem .................................................................................................. 10
   4.2 A Preliminary Note on Why It Matters ..................................................................... 11
   4.3 Identifying A Problem With Kelly’s Argument That Polarized Beliefs Are
      Evidentialist Justified .................................................................................................... 13
   4.4 Unpacking the NDC ................................................................................................ 15
   4.5 Arguing for the NDC ............................................................................................... 19
   4.6 Objections and Replies to Arguments for the NDC ................................................ 22
      4.6.1 The Special Evidence Objection ........................................................................ 22
      4.6.2 The “Better than Chance” Objection ................................................................ 22
      4.6.3 The Accessibility Objection ............................................................................... 25
5. Why It Matters .................................................................................................................. 27
   5.1 Not Our Ordinary Concept ....................................................................................... 27
      5.1.2 The Conservative Picture is Not Suitable as a Revised or a Theoretical
         Account of Justification, Either ............................................................................... 32
   5.2 The Purpose of Our Normative Epistemic Concepts ................................................ 37

**II. Access is Not Necessary for Justification** ................................................................ 46
1. Introduction ....................................................................................................................... 46
   1.1 Theories of Epistemic Justification .......................................................................... 46
   1.2 The Received View ................................................................................................... 48
2. The Motivation for Internalism ....................................................................................... 49
3. Thin Slicing Cases ........................................................................................................... 55
   3.1 The Justification Question ....................................................................................... 57
      3.1.1 Externalism ...................................................................................................... 60
4. Hermeneutic Injustice Cases .......................................................................................... 61
   4.1 The Justification Question ....................................................................................... 65
5. Implications for Our Social Epistemic Practices .......................................................... 67
6. Conclusion and Future Directions .................................................................................. 69

**III. Epistemic Oppression and Ameliorative Epistemology** ......................................... 72
1. Introduction ....................................................................................................................... 72
2. Epistemic Injustice and Epistemic Oppression ............................................................... 73
   2.1 Defining Epistemic Oppression ............................................................................... 77
3. Illustrating the Wrong of Epistemic Oppression ............................................................ 80
   3.1 Fricker’s Testimonial and Hermeneutic Injustice ..................................................... 80
   3.2 Deliberative Injustice .............................................................................................. 83
   3.3 Distributive Epistemic Injustice ............................................................................... 86
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4 Mis-calibrated Intellectual Self-Trust</td>
<td>88</td>
</tr>
<tr>
<td>3.5 Oppressive Normative Epistemic Concepts</td>
<td>91</td>
</tr>
<tr>
<td><strong>4. Epistemic Oppression and Ameliorative Epistemology</strong></td>
<td>94</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td>100</td>
</tr>
</tbody>
</table>
Acknowledgements

I am greatly indebted to all of my advisors, friends, and family for their advice and support, both personal and professional, during the process of writing my dissertation.

Above all, I thank my dissertation advisors, Susanna Siegel, Mark Richard, and Susanna Rinard, for their support, guidance, and feedback through the process. Susanna Siegel, thank you for welcoming me to the department, advising me, and supporting my growth as a philosopher from the very beginning. Your warmth, enthusiasm, and philosophical vigor have been a constant inspiration. Mark, I am ever grateful for your ability to guide me in getting to the heart of big philosophical problems, as well as your patience in poring over the details in order to help me become a clearer and more efficient writer. And thank you for being a firm advocate for myself and so many other graduate students in your role as DGS. We are very lucky to have you. Susanna Rinard, both your careful reading and insightful comments on my work, as well as your astute advice about how to approach and survive both the job market and the profession have been invaluable.

I have regularly presented parts of my dissertation at the Metaphysics and Epistemology Workshop at Harvard, and have consistently benefitted from feedback from both faculty and fellows students there. I also thank audience members from MIT's Epistemology Reading Group, Bled Philosophical Conferences, the Northwestern/Notre Dame Graduate Epistemology Conference, The Senses Graduate Conference at CUNY, the Graduate Conference in Philosophy of Mind, Language, and Cognitive Science at UWO, the Mississippi Philosophical Association Annual Meeting, The Kentucky Philosophical Association Summer Workshop, the Understanding Value Conference at the University of Sheffield, the Northern New England Philosophical Association Conference, the Philosophers' Cocoon Philosophy Conference, and the Society for Women in Philosophy – Ireland's Annual Conference. I am especially grateful to the organizers of those conferences, to my commentators, and to particular audience members who helped me to get the most out of those experiences by offering valuable comments on my work, and on graduate school in general: Terry Horgan, Jack Lyons, Mikkel Gerken, Lauren Ashwell, and Cate Elgin, among others.

The staff of the Department of Philosophy at Harvard have been incredible throughout my time in graduate school. Nyasha Borde, Eric Johnson-DeBaufre, and Ruth Kolodney constantly go above and beyond in their efforts to support graduate students, and I cannot imagine facing the job market without Emily Ware in my corner.

My fellow graduate students have provided a supportive community that has been immensely valuable in bolstering me through the more challenging moments of graduate school. I am especially indebted to the other seven members of my cohort: Alex Prescott-Couch, Byron Davies, Jeremy Fix, Johann Frick, Céline Leboeuf, Elizabeth Miller, and Oded Na'aman; to fellow philosoladies and friends: Olivia Bailey, Ronni Gura Sadovsky, Patricia Marechal, Zeynep Soysal, and Kate Vredenburgh; to James Bondarchuk for his friendship and provision of needed comic relief; to Colin Chamberlain for his constant friendship and for conversations that inspired parts of my writing; to Ryan Doody for helping me think and talk through many of the ideas that ended up in my dissertation; to Zoe Jenkin for excellent comments on an earlier version of my
first paper at the CUNY conference; to Kranti Saran for his generosity in giving me advice as I found my way early in graduate school and beyond; to Ben Bagley and Cat Wade (i.e., the karaoke kweens) for all of the fun we have had together since they joined the department; and most especially, to my fellow philosoladies and core crew, Céline Leboeuf, Elizabeth Miller, and Wendy Salkin, for everything.

I would not have made it through graduate school without the support of my friends and family. I am especially indebted to: Brian Addison, Greg Chalson, Adam Fastman, Jordan Hoffman, Liv Gold, Mandy Hagee, Ian Harper, Tyler Neill, Adam Stack, for keeping me sane; to my siblings: Hilary Morrill and Matt McWilliams, for their support; to my parents Kathy McWilliams and Will McWilliams, for bringing me into existence, raising me, supporting me, and telling me that I could be anything I wanted to be in this world; and of course, to my partner Jack Byrne, for being amazing.
For My Parents
Introduction

My dissertation is in three parts. In each one, I consider phenomena that highlight the moral dimensions of our epistemic lives, and ask what normative concepts best reveal our epistemic situation. I am particularly interested in thinking about how we should define *epistemic justification*. In this introduction, I will say something about the methodology that I use to explore these issues, and I will provide synopses of my main arguments.

My methodology departs from the way that many contemporary analytic epistemologists have theorized about justification and other normative epistemic concepts. Many epistemologists have relied almost exclusively on a kind of conceptual analysis, which begins with the assumption that we have a shared concept, for instance, of *justification*, and seeks to shed light on it by achieving something like a reflective equilibrium between our theoretical commitments and our intuitions about particular cases. The idea is that the principles we are left with at the end of this process give us a more precise picture of our concept.

This is, of course, a useful methodology for discovering the principles that underlie our ordinary thought and talk. But in my view, we ought to do more than that. Epistemology asks normative questions about how human beings should reason and form beliefs. I do not think we can make a good faith effort to answer such questions if we start from the assumption that the answers are already embedded in our commonsense notions. So, the methodology of my dissertation goes beyond conceptual analysis in a couple of ways. First, in theorizing about justification and other normative epistemic concepts that we use to evaluate beliefs, I look to research in psychology for information about the mechanisms that underlie our belief-forming processes. My view is that the science of cognition we accept rightfully constrains our epistemology, because in theorizing about how we should
reason, we ought to consider the powers and limits of the cognitive systems that underlie those reasoning processes.

To give an example: in asking when our beliefs are justified, most epistemic theories rely implicitly or explicitly on a particular understanding of either the belief-forming process or the evidence on which beliefs are based. My methodology is premised on the idea instead of taking these understandings for granted, we ought to look to psychological research to better understand, for instance, what counts as evidence that a subject can use in forming her beliefs.

The further question of how these empirical facts should inform our theorizing brings me to the second way in which my methodology departs from conceptual analysis. Namely, I take an ameliorative approach to theorizing about epistemic concepts. Ameliorative inquiry in epistemology starts by asking what legitimate purposes we might have for employing a framework of normative epistemic concepts to begin with. What work do we want these concepts to do for us in the context of epistemological theorizing? Given an answer to this, we can determine what set of concepts is up to the task. So, the way in which empirical information should inform our theorizing depends on what we should want a concept like justification for to begin with.

In this dissertation, I take an empirically-informed, ameliorative approach to considering various phenomena that highlight the moral dimensions of our epistemic lives, and to asking what they can teach us about our epistemic concepts. In “Epistemic Oppression and Ameliorative Epistemology,” I argue that one desideratum for normative epistemic concepts is to promote our flourishing as epistemic agents. I offer an account of a phenomenon that I call epistemic oppression. It occurs when members of a group are subject to a systematic and unfair disadvantage within a social epistemic structure, resulting from the operation of a capacity to control exercises of their epistemic agency, thus systematically and unfairly limiting their epistemic flourishing. I argue that the concept promotes

---

1 Haslanger (2012).
epistemic flourishing because it allows us to identify and understand the kind of wrong that happens when one is thus unfairly undermined in her capacity as an epistemic agent. In order to think about how to address the kind of wrong that happens when an agent's epistemic flourishing is unfairly impeded, we have to be able to name it, and to explain why it is a problem.

I distinguish several different types of epistemic oppression. One sort of case concerns instances where our normative epistemic concepts themselves can be vehicles of epistemic oppression. For instance, I make a case that there is an oppressive concept of justification embedded in our everyday social epistemic practice. In ordinary interactions, it is common to ask others to cite justificatory reasons for the beliefs they espouse. If a person is unable to do so, then in the context of this practice, we conclude that her belief is unjustified. I argue that this is the wrong result in certain cases where a person is sensitive to features of a situation, but unable to identify or articulate them. I then suggest that in such cases, we need a notion of justification that does not require the person to be able to articulate – or perhaps even access – the justificatory reasons for her belief.

In “Access is Not Necessary for Justification,” I argue that this anti-accessibilist notion of justification also has independent application in cases that are not instances of epistemic oppression. I consider beliefs that are formed on a perceptual basis that psychologists have written about, called thin-slicing. Beliefs based on thin-slicing are rooted in extensive patterns of exposure that are nonetheless unavailable to consciousness. For instance, art historians are able to reliably recognize when a piece of “ancient art” is a forgery, but are unable to identify what those judgments are based on. These cases raise an issue about whether we ought to count these inaccessible perceptual bases for their beliefs as justificatory evidence or not. Certain accessibilist proposals tie the definition of evidence to things that can be arrived at by introspection. That is one conception of evidence, but the phenomenon of thin-slicing presses the issue of whether it is too narrow, by providing a type of real-world case where

---

beliefs are based on inaccessible information in a way that nonetheless makes them reliably true.

I use the ameliorative approach to answer the normative question of whether epistemologists ought to count this inaccessible information as justificatory evidence. In particular, I argue against Michael Bergmann's view that to do so would be to leave ourselves open to a certain powerful objection, and thus give up on the main motivation for being an internalist about justification in the first place. Bergmann suggests that the most compelling reason for endorsing internalism about justification is that if a subject is not aware of what her belief has going for it that distinguishes it from a stray hunch or an arbitrary conviction, then the belief cannot be justified. I argue that when we look more closely at what this suggestion comes to, we see that the issue Bergmann raises should not count as an objection to a theory of justification at all. We can all agree that one thing we want from a theory of justification is to explain how justified beliefs are different from stray hunches and arbitrary convictions. But I argue that under scrutiny, the accessibilist's reason to think that has to come in the form of an awareness of how one's belief is different does not hold up.

In “Evidentialism and Belief Polarization,” I consider a belief revision process often discussed by psychologists, known as belief polarization. It occurs when people who disagree about some complex matter of fact are exposed to a mixed body of evidence that bears on that dispute. Researchers have found that this reliably causes people to increase confidence in whatever their antecedent belief was, such that when subjects start out with different beliefs, the process causes them to polarize with respect to one another. I argue against Tom Kelly's claim that the resulting beliefs are justified on evidentialist grounds.

As in the thin-slicing case, the issue turns on the question of what we ought to count as part of a subject's evidence. Here, the psychological research seems to indicate that in revising their beliefs,

5 Kelly (2008a).
subjects do not respond certain to evidence that is nonetheless *available* to them in a sense, because they are motivated to preserve their antecedent beliefs. I use the ameliorative approach to make a case that given the work we want a theory of justification to do, we ought to hold subjects normatively responsible for that evidence, even when they are not motivated to respond to it. Thus, I claim that *even if* Kelly can carve out an evidentialist concept of justification on which these polarized beliefs come out justified, the resulting notion is not tenable because it fails to give us what we want from a theory of epistemic justification.
1

Evidentialism and Belief Polarization

1. Introduction

Belief polarization occurs when subjects who disagree about some non-straightforward matter of fact are exposed to a mixed body of evidence that bears on that dispute. We might expect that mutual exposure to common evidence would tend to mitigate disagreement, since the evidence available to subjects comes to consist increasingly of items that they have in common. But in fact, the opposite happens: The subjects' disagreement becomes even more pronounced, as each person increases confidence in her antecedent belief. Subjects' beliefs polarize with respect to one another.

Kelly aims to identify the mechanisms that underlie polarization and to assess the normative, epistemic issues that arise from it. In particular, he is interested in whether the processes that underlie polarization undermine the justification of the polarized beliefs. He assumes that subjects' beliefs start justified, and he asks whether these routes to polarization leave the polarized beliefs in good standing. He concludes that given evidentialism, the justification of the subjects' beliefs is not undermined by their being the result of the mechanisms that underlie polarization.

I take on board Kelly's description of the polarizing mechanisms, but challenge his conclusion. I argue that, even by Kelly's evidentialist lights, beliefs that result from these routes to polarization are not justified.

Why does it matter whether evidentialism says that polarized beliefs are justified? Ultimately, I argue that the question is significant because the theory of justification that the evidentialist would need

---

6 Kelly (2008a).
7 Kelly does not explicitly identify himself as a strict evidentialist in his paper. He says only that in paradigmatic cases, how confident it is reasonable to be in one’s belief in some proposition is a matter of how well-supported that proposition is by one’s evidence (Ibid., p. 623). Nonetheless, his argument seems to rely on this being such a case. That is, he argues for his claim that polarized beliefs are rational by appealing to the idea that what it is reasonable to believe is a function of one’s evidence (Ibid., p. 628). So, for our purposes, I will take it that evidentialism is the framework in which Kelly assesses what it is reasonable to believe. Whether he is an evidentialist beyond this will not matter for the issue I am debating here.
to give in order to count these polarized beliefs as justified is out-of-sync with how the term is and ought to be used in ordinary contexts, as well as with how it ought to be used in theoretical contexts. This is a significant conclusion to draw about one of the main contenders for the correct theory of epistemic justification, and the evidentialist will want to avoid this result.

2. Kelly on The Psychological Phenomena That Underwrite Polarization

In identifying the psychological phenomena that underwrite polarization, Kelly draws specifically on an empirical study done by Lord, Ross, and Lepper (1979). They recruited subjects who disagreed about a complex empirical question. In this case, the question was whether capital punishment tends to have a deterrent effect on the commission of murder. Half of the subjects believed $p$: “Capital punishment does have a deterrent effect on the commission of murder”, and the other half believed not-$p$. I'll call these propositions $Deterrent$ and $Not-Deterrent$, respectively.

During the experiment, all subjects were shown the detailed results of two different studies that bore on the disputed question: one study offered support for $Deterrent$, and the other supported $Not-Deterrent$. Subjects were also shown a list of criticisms of each study, and replies to those criticisms. After they had time to reflect on all of this information, subjects on both sides of the issue reported that they had become more confident in their antecedent views about $p$: subjects who antecedently believed $Deterrent$ now felt more confident that it was true, and subjects who believed $Not-Deterrent$ felt more confident in that.

Kelly identifies an empirical phenomenon that underwrites and explains the polarization of subjects' beliefs. I'll call it uneven scrutiny. It refers to the way that Kelly says subjects processed the information in the two studies that they were shown. Namely, they scrutinized the study that disagreed with their view (I'll call this the uncongenial study). That is, they used their cognitive resources to search for flaws that might discredit the study's conclusion: problems with its methodology, variables
that were not adequately controlled for, &c. Meanwhile, when they encountered the congenial study, subjects took its results on board as further evidence for their view, without scrutinizing.

Why did this uneven scrutiny lead to belief polarization? In the Lord, Ross, and Lepper study, both of the studies that each subject was shown contained flaws that would discredit the studies' conclusions. But as a result of uneven scrutiny, the subjects only found the flaws in the uncongenial studies. And because the flaws discredited the studies' conclusions, the subjects took them to be defeaters of the support that the uncongenial study would otherwise lend to the uncongenial view.

To illustrate how this setup results in polarization, consider a subject - call her Elena - who comes into the study believing Deterrent. She is shown a congenial study that supports Deterrent, and an uncongenial one. Since the congenial study agrees with Elena's antecedent belief, she does not scrutinize it, but takes the findings at face value, and increases her confidence in Deterrent. Because the uncongenial study disagrees with her, she scrutinizes it, and discovers the fatal flaws in the study's procedure and methodology. She takes these flaws to defeat the evidence that the study would otherwise lend to Not-Deterrent; so, her credence in Not-Deterrent does not change.8

3. Kelly's Normative Assessment

Kelly is concerned to evaluate the justificatory status of the polarized beliefs that subjects arrive at when they devote more thought to apparent counterevidence in the ways just outlined. His view about what determines the justification of beliefs is a standard kind of evidentialism, which says that whether a belief that \( p \) is justified is (only) a matter of whether \( p \) is sufficiently well-supported by the

---

8 This effect - of prior attitudes leading to uneven scrutiny, and thus to belief polarization - seems robust. It has been documented more recently by Taber & Lodge (2006), in their work on political beliefs.
subject S's evidence. A belief is justified iff it is supported by the evidence that S has. Importantly, this means that how and why S gets her evidence is not (and cannot be) relevant to the justificatory status of her beliefs. So, it does not matter that S gets her evidence from the process of uneven scrutiny just described. For Kelly, S's polarized beliefs are justified because, he thinks, those beliefs are well-supported by S's evidence.

Although he does not say so explicitly, the conception of evidence that Kelly's brand of evidentialism has is mind seems to be, roughly, an accessibilist one. There are many different versions of accessibilism, but in rough terms, it is the view that something counts as part of a subject's evidence only if it is actually or potentially accessible to her by introspection or reflection. For Kelly, access here seems to be roughly synonymous with awareness. So, pieces of evidence are items that the subject can actually or potentially become aware of via introspection or reflection. This definition will suffice for now, though later on I will problematize the question of what exactly accessible evidence comes to for Kelly. Throughout the discussion, I will refer to the collection of things that make up S's accessible evidence as constituting her evidential perspective. With this in mind, Kelly's basic reasoning for the claim that S's polarized beliefs are well supported by her evidence at the end of the study is as follows.

Consider Elena, who starts out believing Deterrent. When she encounters the uncongenial study, she scrutinizes it, and in doing so, gains a defeater for the evidence that it would otherwise give her.

---

9 See, for instance, Kelly (2003). The view - or, at least, its fundamentals - are taken over from what is espoused in Conee and Feldman (2004).
In his (2008a) paper, Kelly makes a distinction between two different senses of ‘evidence’: a broad and a narrow sense. The use of evidence in this formulation of evidentialism refers to evidence in Kelly's broad sense.

10 Strictly speaking, matters may be slightly more complicated, because to secure doxastic (as opposed to propositional) justification, the evidentialist may also want to say that it is necessary that the belief be appropriately based on the relevant evidence. Kelly does not say whether he is talking about doxastic or propositional justification, but it will not matter for the issue we are debating. I think that the polarized beliefs he discusses are doxastically (in addition to propositionally) unjustified, because they are not supported by the evidence that the subject has; so, the question about the basing relation does not arise.

11 There is a further question of whether to interpret this in a way that entails that all pieces of evidence are mental items. Accessibilists would traditionally say that a subject's evidence consists solely of mental items, but there is also room in logical space for the possibility that one actually or potentially could become aware of external items, such as facts about the world, via introspection or reflection. Kelly seems open to this possibility in some of his other writings about evidence. See Kelly (2008b) and Kelly (2008c).

12 A question may arise here about whether, when I say that S gains or gets a defeater, I mean that she discovers a defeater
for Not-Deterrent. Therefore, she does not get new evidence in favor of Not-Deterrent, and her disbelief in Not-Deterrent does not and should not change. When she encounters the congenial study, she does not scrutinize it. She remains unaware of its flaws, so her evidential perspective does not include a defeater for the evidence that the congenial study seems to provide for Deterrent. At the end of the experiment, then, Elena has new information in support of Deterrent, and no new information in support of Not-Deterrent. Therefore, proportioning her belief to the evidence requires her to increase her confidence in Deterrent. When she does, her belief remains justified.

4. Normative (Re)assessment

4.1 Locating the Problem

Intuitively, Elena's polarized belief is unjustified.\(^{13}\) One natural way of vindicating this judgment is by looking to a concept of epistemic (ir)responsibility. You might think that by scrutinizing unevenly, subjects are not being epistemically responsible with respect to ensuring that new sources of evidence are free from fatal flaws. And, you might think a lack of epistemic responsibility during the process of belief formation undermines the justificatory status of resulting beliefs.\(^ {14}\)

I tend towards agreement with the idea that epistemic irresponsibility might affect the justificatory status of resulting beliefs; but I do not think the intuitive point that polarized beliefs are in

---

\(^{13}\) It is an empirical question whether most people find this result to be counterintuitive. For my part, I have noted that amongst the various audiences of undergraduates, graduate students, and professional philosophers that I have polled, it has never not been the case that the vast majority find the result to be counterintuitive.

\(^{14}\) Cf. certain ideas in virtue epistemology

One evidentialist-inspired version of this idea would be that a belief is justified iff it is supported not by the evidence that S actually has, but by the evidence she ought to have, where what evidence she ought to have is determined by looking to our concept of epistemic responsibility.
poor epistemic standing adds much force to arguments for that claim - even granting that the point is true, this is not the only possible explanation of its truth. Perhaps more importantly, an evidentialist like Kelly would not be impressed by this line. Fortunately, there is a more dialectically powerful response available to us. That is, I think it is possible to show on evidentialist grounds that polarized beliefs are not justified.

4.2  A Preliminary Note on Why It Matters

Before I give my assessment of Kelly's evidentialist analysis, allow me to say something preliminary about why it matters whether an evidentialist account of justification counts these polarized beliefs as justified. I will have much more to say about this in the final section of the paper.

Evidentialism is one of the major contenders for a theory of epistemic justification. So it would be worrying if it were committed to the counterintuitive consequence that the polarized beliefs that Kelly describes come out justified. To see how counterintuitive this is, consider, on the concept of justification that counts these polarized beliefs as justified, how little it takes to justifiably become very confident that Deterrent is true. To get there, one would only need to justifiably believe Deterrent to some lesser-but-sufficient degree, and then be exposed to a mixed batch of evidence that bears on whether it is true. The key point, as we saw in §2, is that it does not matter whether the parts of this mixed batch that seem to be in favor of Deterrent are actually any good. Since they accord with the subject's antecedent view, she will take them on board at face value, without scrutiny. And according to Kelly's analysis, the resulting belief is justified. And by iterating this process, one could justifiably become almost certain that Deterrent is true.

Contra Kelly's evaluation, our intuitive judgment that these polarized beliefs are not justified suggests that the concept of justification that counts these polarized beliefs as justified is substantially

15 Recall that for the evidentialist, whether a belief is in good epistemic standing is only a matter of whether it is well-
different from our ordinary one. To make this more vivid, allow me to give another example of the relevant type of polarization process, so that we can imagine it operating outside the context of a belief polarization study. Imagine that you disagree with a peer about a complicated and controversial empirical question. You believe $p$, and she believes $\neg p$. As is typical with complex and controversial questions, there is a lot of information out in the world that seems to bear on the matter. Suppose that in fact, $p$ is true, and that the available information weighs heavily in favor of $p$, in ways that a person of typical competence can appreciate, but only when she puts in the time and effort to properly scrutinize that information. Most of the evidence that seems to support $\neg p$ is flawed in ways that a competent person can recognize when she scrutinizes it. You and your peer are both competent. Further suppose that you believe $p$ because you have spent a lot of time scrutinizing the available information, and you appreciate the ways in which it weighs in favor of $p$. You are confident that $p$ is true, and your belief is justified. Your peer believes $\neg p$ because she happens only to have come across a few bits of information that bear on whether $p$, and they all happen to support $\neg p$, in ways she comes to appreciate by scrutinizing them. Since she recognizes that the matter is complex, and knows that she has not put a lot of effort into its consideration, her belief that $\neg p$ is not very confident. By evidentialist standards, her belief is also justified. So far, there is nothing counterintuitive about this. The next bit is where things get tricky.

One day, your peer becomes convinced that this matter is an important one, and thus motivated to form a considered view. She goes out into the world to gather and consider a substantial, mixed batch of evidence that bears on whether $p$. But because she antecedently believes $\neg p$, she subjects that evidence to uneven scrutiny, in the way that Elena did. In this case, this means that although much of the evidence that seems to support $\neg p$ is flawed, she takes it on board without scrutiny. At the end of this process, she believes $\neg p$ with the utmost confidence.\footnote{For those readers who prefer to fill in this schema with a real life example: the literature suggests that belief polarization is prone to happen with beliefs about climate change. Interestingly, a recent study found that polarization happens most}
According to the evidentialist analysis that Kelly endorses, her belief is justified. But here, our ordinary judgment about the case says otherwise. This suggests that our ordinary concept of justification is not one that accommodates your peer's belief in this scenario. And given the features of the case, that should not be terribly surprising. In ordinary contexts, when we judge that another person's belief is justified, that amounts to a kind of endorsement. And granting such endorsement in incompatible – particularly in a case where a person believes a proposition with the utmost confidence – with knowing that the belief is based on a shallow consideration of bad evidence. So, the picture of evidentialist justification on which Elena's belief and your peer's belief count as justified is out-of-sync with the ordinary concept that governs our intuitive judgments in these contexts. Ceteris paribus, this is reason to worry about this concept of justification. In the final section of the paper, I will make the stronger argument that there is more than ceteris paribus reason to worry about it.

4.3 Identifying A Problem With Kelly's Argument That Polarized Beliefs Are Evidentialist Justified

Kelly’s evidentialism implies that the polarized beliefs at the end of the Lord, Ross, and Lepper study are unjustified if in acquiring them believers fail to proportion their belief to the new evidence that they get during the course of that study. I will argue that this is just what happened. The problem is not with the way that subjects responded to evidence from the uncongenial study, but rather with the way that they responded to the results of the congenial study, by taking them on board at face value. When subjects take the results of the congenial study at face value and increase their confidence, the resulting polarized belief are not well-supported by their total evidence. Thus, they are unjustified on evidentialist grounds.

---

strongly among people with the highest degrees of scientific literacy and technical reasoning capacities. A plausible interpretation of this research is that those who possess these technical competencies are all the more successful at scrutinizing evidence that seems to disagree with their view, in order to find flaws that allow them to dismiss that evidence.

See Kahan et al. (2012).
To show this, I will reconstruct Kelly's argument that polarized beliefs remain well-supported by the subject’s total evidence, and I will say where I think that argument goes wrong, with the ultimate result being that polarized beliefs are not well-supported by their total evidence. Recall that on Kelly's analysis of the case, during the course of the experiment, Elena, who antecedently believes *Deterrent*, gets the following new evidence:

E1: [The uncongenial study presents a set of empirical results that seems to support *Not-Deterrent*], [The study's authors claim that those results support *Not-Deterrent*]

E2: [The congenial study presents a set of empirical results that seem to support *Deterrent*], [The study's authors claim that those results support *Deterrent*]

U1: [The uncongenial study contains an undercutting defeater of the evidence for *Not-Deterrent* in E1] 17

So, according to Kelly, Elena's total (new) evidence at the end of the study is: {E1, E2, U1}. And U1 defeats E1. Kelly thinks that if Elena’s total evidence is her old evidence plus these three items, with the first supporting *Deterrent*, and the rest not supporting it, then her total evidence comes out more strongly in favor or *Deterrent*. So, she should become more confident. More specifically, his argument is this 18:

(P0) Elena starts out with evidence for *Deterrent* (assumption).

(P1) The new evidence that Elena gets during the experiment is (E1, E2, and U1).

(P2) E2 supports *Deterrent*.

(P3) E1 and U1 do not support *Not-Deterrent*.

Therefore, [from (P0), (P1), (P2), and (P3)]

(C1) Elena's total evidence supports *Deterrent*.

17 In the next section, I will explain what an *undercutting defeater* is, and will make a further distinction that helps us narrow in on the type of defeater we have in this case.

18 Kelly does not put his argument in premise-conclusion form; this is a reconstruction.
(P4) Evidentialism: You should proportion your confidence to what your total evidence supports.

Therefore, [from (C1) and (P4)]

(C2) Elena should become more confident in Deterrent.

My main criticism of this argument is that, though seductive, the inference from [(P0), (P1), (P2), and (P3)] to (C1) does not go through. This, I claim, is because we cannot “read off” a fact about what Elena's total evidence supports from facts about what individual pieces of it support. The individual pieces of evidence may bear on each other in ways that are not captured by looking at each piece in isolation.

In this case, I think that what Kelly overlooks by looking at each piece of evidence in isolation is that U1 is a normative, undercutting defeater of the evidence that Elena gets in E2. I'll call this claim - that U1 is a normative, undercutting defeater of the evidence that she gets in E2 - the Normative Defeater Claim (NDC). More specifically, the claim is that U1 adds something to Elena's evidential perspective that undermines the evidential support relation between E2 and the truth of Deterrent to a significant degree, if not completely.

4.4 Unpacking the NDC

The main thrust of the NDC is that there is a relationship of defeat between U1 and E2, in addition to that between U1 and E1. Both are relationships of undercutting defeat. The difference is that in the U1/E1 case, U1 is both a psychological and a normative defeater, while in the U1/E2 case, it is only a normative defeater. Let me explain.

An undercutting defeater is one that undermines the evidential support relation between one's
evidence for a proposition, and the truth of that proposition.\textsuperscript{19} The evidence thus loses its status as evidence for that proposition. In our case, U1 undermines the evidential support relation between E1 and Not-Deterrent. Or (equivalently), U1 undermines E1's status as evidence for Not-Deterrent. In this case, Elena grasps U1 when she scrutinizes the uncongenial study, and finds that the justification for the study's claim that its empirical results support Not-Deterrent was fatally flawed, such that their having seemed to support Not-Deterrent was merely illusory. Another way of putting this is that for Elena, U1 acts as a \textit{psychological} as well as a \textit{normative} defeater of E1.

The distinction between psychological and normative defeaters is owing to Jennifer Lackey, though I am not certain whether I am using the terms in exactly the sense she does.\textsuperscript{20} I take a \textit{psychological} defeater to be one that acts as a defeater in the subject's psychology, such that the defeated evidence loses its status as evidence for the relevant proposition from the subject's perspective. For instance, if S became more confident in a proposition \( p \) on the basis of evidence that was later psychologically defeated, her confidence would then return to its antecedent level. A \textit{normative} defeater is something within S's evidential perspective that \textit{ought} to act as a defeater in her psychology, but may not. The \textit{ought} here is the \textit{ought} of epistemic rationality: when a normative defeater does not also act as a psychological defeater, this amounts to a failure on the part of the subject to proportion her beliefs to her evidence. The idea that underlies both psychological and normative defeaters, as Lackey puts it, is that certain kinds of experiences, doubts, and beliefs contribute epistemically unacceptable irrationality to doxastic systems. Both psychological and normative defeaters, as Lackey puts it, is that certain kinds of experiences, doubts, and beliefs contribute epistemically unacceptable irrationality to doxastic systems. Both psychological and normative defeaters

\textsuperscript{19} This definition is due to Pollock (1986).
\textsuperscript{20} Lackey puts the distinction this way: “A psychological defeater is an experience, doubt, or belief that is had by S, yet indicates that S's belief that \( p \) is either false or unreliably formed or sustained. Defeaters in this sense function by virtue of being had by S, regardless of their truth value or justificatory status. Second, there are what we might call \textit{normative defeaters}. A normative defeater is a doubt or belief that S ought to have, yet indicates that S's belief that \( p \) is either false or unreliably formed or sustained. Defeaters in this sense function by virtue of being doubts or beliefs that S \textit{should have} (whether or not S does have them) given the presence of certain available evidence.” Lackey (2006), p.4. I am not certain whether Lackey's normative defeaters would include doubts and beliefs that S ought to have in the sense that she ought to go out into the world and acquire further evidence that bears on the veracity of her evidence for \( p \). Kelly's breed of evidentialism would not accept the existence of such normative defeaters.
defeaters are defeaters in this sense, but only psychological defeaters, by definition, play the role that they epistemically ought to play in the subject's psychology.

The *NDC* says that U1 is a normative, undercutting defeater of E2. Having introduced some terminology, we can say that U1 is a normative but not a psychological defeater of E2. Elena does not take herself to have a defeater for E2. Rather, she treats E2 as evidence for *Deterrent*, and boosts her credence accordingly. Per our definition, to show that U1 is a normative defeater of E2, I will argue that the fact that Elena does not understand or treat U1 as an undercutting defeater of E2 amounts to a failure on her part to proportion her belief to her accessible evidence.

Finally, to understand the *NDC*, I need to say something about the content of E2: [The congenial study presents a set of empirical results that seem to support *Deterrent*], [The study's authors claim that those results support *Deterrent*]. Having a better grasp of this content will allow us to understand what it would take for E2 to get defeated. Both components of E2 involve the proposition that the congenial study's empirical results support *Deterrent*, where those results are a set of rather straightforward empirical findings, and the idea that they support *Deterrent* is the idea that the fact that the findings were such and such lends credence to the truth of *Deterrent*. To make clear how this works, let's look at an actual example of results that subjects were shown during the Lord, Ross, and Lepper study. The following is one set of results that subjects were presented with.

Kroner and Phillips compared murder rates for the year before and the year after adoption of capital punishment in 14 states. In 11 of the 14 states, murder rates were lower after adoption of the death penalty. This research supports the deterrent effect of the death penalty.²¹

Note that this statement conveys some straightforward empirical results of the study's investigation, and then suggests that those results support the truth of *Deterrent*. In other words, participants get a

²¹ Lord, Ross, & Lepper (1979).
statement of the form: “Empirical results $R$. $R$ supports Deterrent.” Here, the straightforward empirical results in $R$ are that Kroner and Phillips compared murder rates for the years described in the states described, and that in 11 of 14, those rates were lower after adoption of the death penalty.\(^{22}\) I think it is epistemically okay for Elena to accept $R$ at face value. The study's authors are not likely to be wrong about $R$. The further proposition, that $R$ supports Deterrent, is the tricky part. And E2 is made up of the fact that the study's authors claim that $R$ supports Deterrent, and that it seems to Elena that $R$ supports Deterrent. So, what do the congenial study's claim that $R$ supports Deterrent, and the seeming that $R$ supports Deterrent, amount to? Support is something that can come in degrees. Support is also relative to an evidential perspective. The congenial study's claim that $R$ supports Deterrent means that from its authors' evidential perspective, the truth of Deterrent is likely to be what accounts for $R$, or at least to be a substantial contributor.\(^{23}\) Of course, whether the authors have properly taken account of everything that is in their evidential perspective is a further matter, but the act of making a claim that $R$ supports Deterrent in this context seems to at least involve purporting to have taken proper account of all available evidence. The other part of E2 – its seeming to Elena that $R$ supports Deterrent – is perhaps more inchoate. It means that from her evidential perspective, the truth of Deterrent seems likely to be what accounts for $R$ (or at least a substantial contributor). But being a seeming and not a claim, it does not involve a purporting on Elena's part to have taken account of everything that is within her evidential perspective.

On the surface of it, it certainly seems like Deterrent is the sort of thing that is likely to account for $R$. But of course, other factors could also explain $R$, such as confounding variables, methodological

---

22 More specifically, let $R$ stand for the fact that the researchers followed this procedure, and got this result.

23 To put this more formally, in terms of Bayesian confirmation theory: $R$ supports Deterrent just in case, given everything that is within Elena's evidential perspective, the prior probability of Deterrent conditional on $R$ is greater than the prior unconditional probability of Deterrent. What I have articulated is a description of what would make this true in our particular case.

For further discussion from the Bayesian perspective, see Earman (1991) and Fitelson (1999). But I think that for our purposes, the intuitive description I’ve given should be sufficient.
issues, and the like. What determines the extent to which \( R \) supports \textit{Deterrent} from a subject's evidential perspective is the degree to which they have reason to think other factors could plausibly contribute to explaining \( R \) (and, conversely, the degree to which they have positive reason to think there are \textit{no} other factors that might explain \( R \)). In general terms: \( R \) lends \textit{some} credence to \textit{Deterrent} so long as, given a subject's evidence, \textit{Deterrent} is likely to be what accounts for \( R \) (or at least to make a substantial contribution).

Defeat of \( E_2 \) can come in degrees, too. To the extent that Elena gains access either – (1) to evidence indicating that \( R \)'s having seemed to her to support \textit{Deterrent} was merely illusory, and not caused by the corresponding set of facts out in the world; or (2) to evidence that the congenial study's authors may have been wrong to claim that \( R \) supports \textit{Deterrent} – the evidential support relation between those parts of \( E_2 \) and the truth of \textit{Deterrent} is undermined.\footnote{That is, those bits of evidence are defeated. Here I have not distinguished between psychological and normative defeat.} The \textit{NDC} says that when Elena uncovers \( U_1 \), this adds something to her evidential perspective that thus undermines the evidential support relation between \( E_2 \) and the truth of \textit{Deterrent}. More specifically, given that defeat can come in degrees, the claim is that \( E_2 \) is defeated to some significant degree, if not completely. If Elena does not adjust her confidence accordingly, then her belief is in poor epistemic standing on evidentialist grounds. Having thus unpacked the \textit{NDC}, I will now argue for it.

\textbf{4.5 Arguing for the NDC}

In arguing for the \textit{NDC}, I am going to assume that Elena sees the congenial study before the uncongenial one,\footnote{In the Lord, Ross and Lepper (1979) study, half of the subjects see the congenial study before the uncongenial one, and half see the uncongenial one before the congenial one.} because I think this is the harder case to argue for (since in this case, she is getting a retrospective defeater of evidence that she has already seen, as opposed to already having the defeater before her mind when she encounters the evidence).
So: Elena first looks at the congenial study, takes it at face value, and boosts her confidence in *Deterrent*. Then she looks at the uncongenial study, scrutinizes it, and finds fatal flaws. In doing so, she discovers that it is what I will call an *evidential dud*. Being an *evidential dud* means that neither the fact that the authors claim that “*R supports p*”, nor that it seems to the subject that *R* supports *p* gives her *any* evidence in favor of *p*. In our case, it means that neither component of E1 gives Elena any evidence in favor of *Not-Deterrent*. She now knows that the seeming was illusory, and the authors' reasoning for their claim was fatally flawed. So, upon discovering U1, Elena comes to know that the uncongenial study is an *evidential dud*.

The *NDC* claims that U1 acts as a normative defeater of E2 as well. So, when Elena discovers U1, it should retrospectively defeat (at least most of) the credence that E2 lends to *Deterrent*. This is true for two reasons.

First: understanding how U1 defeats E1 adds something to Elena's evidential perspective such that if she is proportioning her beliefs to her evidence, she will become skeptical of the congenial study's claim that its results support *Deterrent*. To see this, recall that both the uncongenial and the congenial study's results are of the form “*R. R supports p*”,26 where *R* is the kind of thing that *could* explain *p*, and that *prima facie* seems to support *p*. When Elena discovers U1, this adds more than the bare *fact* of U1 to her evidential perspective. She now has an example and an understanding of a way in which studies that make *prima facie* plausible claims of the form “*R supports p*” can turn out to be evidential duds. Furthermore, she gains an experiential understanding that the kinds of fatal flaws that confer dud-status can be non-obvious, since she had to scrutinize in order to find U1. The result of adding all of this to her evidential perspective is that proportioning her beliefs to the evidence now requires Elena to revise her stance towards the congenial study's claim, and the seeming that its results support *Deterrent*. Rather than taking these at face value, evidential rationality now requires that she be

---

26 Here, of course, *p* represents *Deterrent* and *Not-Deterrent*, respectively.
skeptical. U1 acts as a normative defeater of Elena's warrant for E2. Accordingly, epistemic rationality requires that she lower her credence in *Deterrent*, since it has just been bolstered by taking the congenial study's claim at face value.\(^{27}\) If she does not, then her belief is unjustified on evidentialist grounds.

The second reason that U1 acts as a normative defeater of (at least most of) the credence that E2 lends to *Deterrent* is that Elena's discovery that the uncongenial study is an evidential dud should raise her credence that the congenial study is a dud, too.

Suppose Elena came into the Lord, Ross, and Lepper study with a justified belief that it is very rare for these types of empirical studies to be evidential duds (in Bayesian terms: her prior probability distribution for how often studies of this sort are evidential duds is skewed heavily towards its being very rare). This is a charitable assumption, because if she started off believing duds are somewhat common, then it was irrational in the first place for her to take E2 at face value and increase her credence in *Deterrent* without first scrutinizing the congenial study to rule out the possibility of its being a dud.

After she scrutinizes the uncongenial study, she finds that it *is* a dud. At that point, the fact that at least one of the two studies she has been shown is an evidential dud becomes part of her total evidence. Importantly, this should lower her confidence that evidential duds are quite rare, because evidential rationality requires her to conditionalize on this new knowledge that one of the two studies

\(^{27}\) You might think that Elena is not aware that she became more confident in *Deterrent* in response to taking this claim at face value, so that fact is not part of her evidence, so she does not gain a reason to correct this mistake by lowering her credence. But we know that at the very least, subjects are aware of having become more confident in response to their exposure to this evidence, because they self-report their increased confidence at the end of the experiment. Even if she is not aware of the specifics, it is accessible to her that reading the congenial study made her feel more confident. So if she now has reason to be skeptical of that study, proportioning her belief to that new evidence requires a decrease in confidence.
she was shown is a dud. And this should affect what she thinks about whether the congenial study is a dud. Since she now believes that there are more duds out there than she previously thought, she ought to boost her confidence in the proposition that any given study – including the congenial one – is a dud.

**4.6 Objections and Replies to the Arguments for the NDC**

**4.6.1 The Special Evidence Objection:**

In response to the foregoing argument that Elena should raise her credence that the congenial study is a dud, one might object that Elena has some evidence that the congenial study in particular is not a dud. Namely: she read it, and did not find any fatal flaws.

But this should not count for Elena as evidence that the congenial study is not a dud. Recall: it was not until she scrutinized the uncongenial one that she was able to find the fatal flaw, and thus discover that the study was a dud. She therefore sees that the kinds of fatal flaws that confer dud-status can require effortful scrutiny to uncover. So the impression that she got after reading the congenial study – that it seemed plausible – is not good evidence that it did not contain a fatal flaw. Since she did not scrutinize the congenial study, she is not in a good evidential position to judge whether it has such a flaw.

**4.6.2 The “Better than Chance” Objection:**

My second point in favor of the NDC was that when Elena discovers that at least one of the two

---

28 You might think that whether she has access to this fact depends on what access come to. Perhaps the subject's psychology is compartmentalized in a way that makes it hard for her to put two and two together. I will talk at greater length in §4.6.3, and again in §5, about what access comes to. But a very basic preview of my reply is that when subjects are motivated to form true beliefs about the matter at hand, they are able to put two and two together, and in light of this, we ought to hold epistemic agents to this standard.

22
studies she has been shown is a dud, evidential rationality requires her to conditionalize on this new
evidence, thus lowering her credence that duds are very rare, and boosting her credence that the
congenial study is a dud. But how much she ought to boost it depends on a few different factors. For
one, it depends on how confident she was in her initial belief that it is very rare for these types of
studies to be duds. For another, it depends on how we think she ought to understand the relevant class
of these types of studies. You might think she ought to conceive of these types of studies very broadly,
as any that aim at settling a complex empirical matter using relevant data. At the opposite extreme, you
might think that since she initially believed duds were very rare, her discovery that one of the studies is
a dud is best explained by the idea that the studies she is being shown in this context are not a random
sampling from the relevant class, so that in this context, she ought to understand these types of studies
as studies that I am being shown in the context of this experiment.

So, how much Elena is rationally required to lower her credence that duds are very rare depends
on a number of things. The objector might point out that on many reasonable ways of filling this in, it
turns out that Elena is still rationally permitted to believe that duds are somewhat rare. And so long as
she is rationally permitted to believe that more than 50% of these studies are non-duds, she may end up
with a justified polarized belief. If she justifiedly believes that more than 50% are non-duds, then even
if she has no special reason to believe that the congenial study in particular is a non-dud, she may still
justifiedly believe it is more likely than chance to be a non-dud. And if that is right, then she still ought
to raise her credence in Deterrent to some degree, relative to where it was when she came in. In other
words: conditional on there being a study that has a greater-than-50% chance of containing genuine
evidence of the truth of Deterrent, credence in Deterrent should go up, even if only a little. So, the
objecor says, it is still the case that when Elena leaves the study, her credence in Deterrent ought to be higher than it was when she came in. This would mean that her belief is both polarized, and epistemically rational.

This seems right, so far as it goes. For the reasons the objector describes, it may well turn out that at the end of the study, Elena’s credence in Deterrent ought to be slightly higher than it was when she came in. Whether it in fact turns out this way will depend on the factors I mentioned earlier, as well as on how reliable she thinks non-duds are (since studies that lack fatal flaws may still not be 100% reliable). But more importantly, this route to polarization is different from the one that Kelly describes as happening in the Lord, Ross, and Lepper study. When Elena boosts her credence in Deterrent, she is not responding to the fact that given her belief about the frequency of duds amongst this type of study, it is more likely than chance to be a non-dud, and thus more likely than not to contain some genuine indication of the truth of Deterrent. Rather, she is feeling the force of not having found anything wrong with the congenial study, and thus continuing to treat that as evidence that it is a non-dud. This is evinced by the fact that when asked, subjects expressed the belief that the congenial study was convincing.29 They did not express the belief that given their total evidence, the congenial study had a greater than chance probability of being a non-dud, and therefore provided some evidence in favor of their antecedent belief, even without scrutinizing it to see whether this probable state of affairs matches the actual one. So, my claim that the route to polarization that Kelly describes leaves beliefs in poor epistemic standing survives the objection. It may be that there is an evidentialist-rational route to polarization, but it is not the one he describes.

29 Of the congenial study, subjects said things like, “It shows a good direct comparison between contrasting death penalty effectiveness. Using neighboring states helps to make the experiment more accurate by using similar locations” and “It does support capitol punishment in that it presents facts showing that there is a deterrent effect and seems to have gathered data properly". Quotes like this, I take it, express subjects’ beliefs that they do have special reason to think that the congenial study is a non-dud.
4.6.3. The Accessibility Objection:

One might resist the claim that U1 is a normative defeater of E2 by questioning whether that defeat relation is really accessible to the subject. In particular, the question would be: is U1 accessible to Elena qua undercutting defeater of the support that E2 lends to Deterrent? You might think not. You might think, for instance, that Elena's beliefs are compartmentalized in a way that would prevent her, even upon active introspection and reflection, from becoming aware of the defeat relationship between U1 and E2.  

Certainly, I grant the possibility that an epistemic agent could be such that her beliefs are compartmentalized in this way. But I think there is evidence in the psychological literature that we are not generally like this. Instead, the evidence seems to indicate that the problem is just that Elena is not presently motivated to find problems with information that seems to support Deterrent. And there is evidence that her motivation could change such that it would be just as easy for her to see the defeat relation between U1 and E2 as it is for her to see that between U1 and E1.

In the psychological literature on attitude polarization and motivated reasoning, a distinction is made between two types of motivation; namely, defense motivation and accuracy motivation. A subject is said to be defense motivated when her primary motivation is to vindicate an existing belief. She is said to be accuracy motivated when her primary motivation is to figure out the truth about some matter, even if the truth turns out to be contrary to her existing belief about it. In our case, it seems clear that subjects in the Lord, Ross, and Lepper (1979) study are defense motivated. This is why they...

---

30 Thanks to Zoe Jenkin for pointing this out to me.

31 Granting this possibility may mean that I have to bite the bullet and concede that this hypothetical agent's polarized belief would count as evidentialist-justified. Part of the question concerns just how much unsuccessful introspection or reflection is allowed before the agent counts as “not having access” to the defeat relation. Some accessibilists (see Ginet, 1975, p.34; quoted in Alston, 1989, p.213) insist that the evidence has to be “directly recognizable” in order to count as accessible, but one need not go this way. Supposing the defeat relation is not accessible on any amount of introspection and reflection, I have to concede that the agent's belief is evidentialist-justified. If you judge that this is the wrong result, then perhaps it is so much the worse for evidentialism. Nonetheless, I do not think the subjects that Kelly has in mind are generally like this.

32 For discussion of this distinction, see, for instance, Chaiken et al. (1996); Chen & Chaiken (1999); Ditto et al. (1998); Gawronski & Bodenhausen (2006); Hart et al. (2009).
scrutinize unevenly in the way I have described. But there is ample evidence in the psychology literature that the relative strength of subjects' defense and accuracy motivations can be manipulated rather easily with experimental interventions, and that when subjects are accuracy motivated, they do not engage in uneven scrutiny. Thus, when subjects are accuracy motivated, their beliefs do not polarize after exposure to the kind of mixed information that is given to them in the Lord, Ross, and Lepper study.

I predict that if, instead of asking subjects to identify their pre-existing belief at the start of the experiment (which increases defense motivation by giving them a feeling of having a stake in the game), Lord, Ross, and Lepper had simply offered their subjects a reward for forming an accurate belief about the matter at hand (thus inducing a stronger accuracy motivation), the subjects would have scrutinized evenly. Furthermore, I predict that under these conditions, when she discovered U1, Elena would have recognized that it bears a relation of undercutting defeat to E2, for the reasons I described in the previous section.33

The objector might push back by insisting that these predictions about what is accessible in a hypothetical situation do not tell us anything about what is accessible to Elena in her actual situation. Perhaps, given her present defense motivation, her beliefs really are compartmentalized to the extent that objector suggests. After all, we determine what is accessible by looking to what the subject can access solely by introspection or reflection, not by looking to what she could access by introspection or reflection, given a certain kind of change in her motivational state.

The question comes down to what the ability to access the defeat relation between U1 and E2 consists in. On my picture of what it consists in, Elena has the ability, and simply lacks the proclivity. In a similar sense, my dog has the ability to jump over the 3-foot fence that encloses our yard, but she generally lacks the proclivity to do so unless there is something enticing on the other side (a rabbit will

33 Ultimately, of course, this is an empirical claim. But I think it is vindicated by looking to the discussions in the psychological literature mentioned in the previous footnote.
do the trick, for instance).

There are, of course, further questions about whether this is the right way to understand the kind of ability to access something by introspection or reflection that is relevant for an evidentialist theory of justification. In the next and final section of the paper, I will return to the question of why it matters whether an evidentialist theory of justification yields the result that polarized beliefs are justified. Along the way, I will make a case that the evidentialist may have reason to understand access in the way I have suggested here, rather than in a more restrictive manner that would exclude the defeat relation between $U_1$ and $E_2$ from counting as accessible.

5. Why It Matters

5.1 Not Our Ordinary Concept

Finally, we are in a position to re-visit the question of why it matters whether an evidentialist theory of justification counts the polarized beliefs that Kelly discusses as justified. In §4.2, I said that a concept that counts these beliefs as justified is worryingly out-of-sync with our ordinary concept of justification. If the epistemologist's aim in theorizing about justification is to capture ordinary use, then this presents a problem for any version of evidentialism on which these polarized beliefs come out justified. But sometimes, the aim of philosophical theorizing is not to capture ordinary use. Sometimes philosophical analyses depart from our ordinary concepts, either to revise and improve them, or in the service of developing independent, theoretical concepts that do not relate to our ordinary ones in a straightforward or obvious way. Perhaps Kelly's thought is that evidentialism offers such a revised or theoretical analysis of justification, which accommodates the polarized beliefs he writes about. If so, then it would be less worrying that the account of justification that he has in mind is out-of-sync with our ordinary concept of justification.
Ultimately, I argue that the evidentialist concept of justification that accommodates polarized beliefs like Elena's cannot successfully offer a revised or a theoretical account of justification, either. In what follows, I look at what the evidentialist's concept of justification would have to be like in order for polarized beliefs to count as justified. I argue that in addition to being out-of-sync with the concept that terms like justification pick out in ordinary contexts, the concept of justification that accommodates polarized beliefs is out-of-sync with the concept they ought to pick out in both ordinary and theoretical ones.  

What would Kelly's evidentialist concept of justification have to be like in order for the polarized beliefs he writes about to count as justified? My argument that Elena's belief is unjustified on evidentialist grounds is based on the idea that the defeat relation between U1 and E2 counts as part of her accessible evidence. I argued that Elena has the ability – though she lacks the proclivity – to access this defeat relation. But perhaps, for Kelly, not everything that a subject has the ability to access counts as accessible. If so, then perhaps for him, the defeat relation between U1 and E2 is not part of the accessible evidence to which Elena must proportion her belief in order for it to count as justified.

Kelly is not explicit about what counts as accessible evidence, even though his argument seems to turn on how we define it. In general terms, what the accessibilist picture usually comes to is that a person's accessible evidence is everything she has actual or potential access to by introspection or reflection. But there are different ways of spelling out what is meant by potential access, some of which are more restrictive than others. A particularly restrictive option might say that a subject has potential access to a piece of evidence only if she can access it by simply turning inward and observing what is immediately given to her, so to speak. A less restrictive one might count a piece of evidence as potentially-accessible even if it takes a significant amount of cognitive effort to access it by

---

34 When I refer to “terms like justification” here, I of course have in mind the terms justified and unjustified. But I would also include terms like rational, irrational, reasonable, and unreasonable, which are often used interchangeably with justified and unjustified to evaluate beliefs in ordinary contexts.
introspection or reflection. Since the defeat relationship between U1 and E2 may take at least some active cognizing to access by reflection, it would not count as potentially accessible on a maximally restrictive construal of what that comes to.

In order for his argument that Elena's belief is justified to go through, then, Kelly would need a fairly restrictive view about what counts as accessible evidence. Perhaps he would be happy to accept this. After all, on his view of these matters, there is a fundamental distinction between epistemic and instrumental rationality, such that questions about how much effort one should devote to scrutinizing a given piece of evidence are practical questions.\(^{35}\) So it would not be altogether surprising if he treated the question of how much \textit{cognitive} effort to put into accessing one's evidence as a practical one, too, such that one cannot have epistemic reason to exert such effort in order to access evidence by introspection or reflection. If so, then for Kelly the evidentialist mandate to proportion one's belief to her total accessible evidence may only refer to evidence that is easily and unproblematically accessible to the subject when she turns inward via introspection or reflection.\(^{36}\) In rough terms, we can think of this as a \textit{perceptual} model of introspection and reflection: when a subject turns inward, certain evidence is available without the exertion of cognitive effort in the same way that when she opens her eyes, the way that things look to her is automatically available.

If we plug this construal of what counts as accessible evidence into Kelly's version of evidentialism about justification, then his argument that Elena's belief is justified may well go through. But it comes at a cost. Not only is the resulting evidentialist view of justification out-of-sync with the concept that the term \textit{justification} refers to in ordinary contexts, but I will argue that it is also out-of-sync with the concept that it \textit{ought} to refer to in both ordinary and theoretical ones. The upshot is that it results in a theory of justification that is untenable across the board. In what follows, I begin by looking

\(^{36}\) I will set aside the question of how much cognitive effort might be considered too much on this model – it does not matter, so long as the defeat relation between U1 and E2 is beyond the pale.
the evidentialist notion that Kelly would need for these polarized beliefs to come out justified, and considering how it is different from our ordinary concept of justification. I then draw on this in explaining how and why that evidentialist concept is different from what both our ordinary and theoretical concepts of justification ought to be like.

Evidentialism about justification begins with the idea that a subject's belief is justified iff it is proportioned to her total accessible evidence. Earlier we saw that for Kelly, this means that why and how a subject gets her evidence is not epistemically relevant; all that matters is that she proportions her belief to the evidence that she ends up with. So it is not relevant that in Elena's case, she could have had different evidence if she had processed the information in the congenial study in a less biased manner. Kelly's view does not leave room for holding a subject accountable for this aspect of her belief revision process. And, we have just seen that when we plug in the restrictive view of access that a positive evaluation of Elena's belief requires, the result is that the subject is not held epistemically accountable for evidence that takes cognitive effort to access by introspection or reflection, either. The upshot is that we are left with a picture of justification on which a subject's belief cannot be criticized either for (i) her failing to proportion her belief to evidence that she could have access to if she had processed information in a less biased manner; or for, (ii) her failing to proportion her belief to evidence that she could access upon further reflection – such as the defeat relation between U1 and E2. Call this picture on which (i) and (ii) are not epistemically relevant the conservative picture of evidentialist justification. The conservative picture is the one that Kelly needs in order to get the result that Elena's belief is evidentialist justified.

On our ordinary notion of justification, (i) and (ii) seem like the sort of things that might well undermine the justificatory status of a subject's belief. This is suggested by the fact that these seem to articulate some of the things that drive our intuitive judgment that Elena's belief is unjustified. To see this, imagine a slightly different version of the case, starring Elena2. Elena2 gets exactly the same new
evidence that Kelly says Elena has: \{E_1, E_2, U_1\}. But Elena2 acquires the evidence via testimony. Her version of the Lord, Ross, and Lepper study is set up so that someone else reads all of the information about the congenial and uncongenial studies, and then conveys it to her. And Elena2 is given ample reason to believe that her informant is not only maximally competent with respect to interpreting these types of studies, but also completely unbiased, and epistemically virtuous in every possible respect. Unbeknownst to Elena2, however, her informant turns out to be biased, and to fall short of her potential to be an exemplar of competence and epistemic virtue in this particular case. Indeed, the informant processes the information in the exact same manner as Elena did in the original case, subjecting it to uneven scrutiny, and failing to reflect on how the different parts of her evidence bear on one another.

Intuitively, when Elena2 ends up with the same polarized belief as Elena did in the original case, her belief is justified. The difference between the two cases is that Elena2's setup removes the possibility of criticizing her belief on the basis of (i) and (ii). Regarding (i): since Elena2 bears no responsibility for the genesis of her evidence, there is no question of evidence she could have had access to if she had processed information in a less biased manner. As for (ii): Elena2 in effect has a defeater defeater for the defeat relationship between U_1 and E_2, since she has every reason to think that her informant is maximally competent and unbiased; that is, she has reason to think that if the congenial study contained a fatal flaw, the informant would have found it and reported it to her. When we remove the possibility of criticizing the subject's belief on the basis of (i) and (ii), but keep her evidence the same, we are no longer inclined to judge that her polarized belief is unjustified. This suggests that on our ordinary concept of justification, (i) and (ii) articulate some of the very things that can undermine the justificatory status of a subject's belief.

So, for Kelly's argument that Elena's belief is justified to go through, he needs a conservative picture of evidentialist justification, on which a belief's justificatory status cannot be undermined by (i) or (ii). But (i) and (ii) seem to capture some of the very features our intuitive judgment responds to.
when we feel that Elena's belief is unjustified. This suggests that, at least at the level of what these judgments respond to, the conservative picture is quite different from our ordinary concept of justification.

5.1.2 The Conservative Picture is Not Suitable as a Revised or a Theoretical Account of Justification, Either.

I will argue that the conservative picture of evidentialist justification does not give us what we should want from a revisionary or a theoretical account of justification, for a couple of reasons. The first is simple. It is not suitable as a revision of our ordinary notion because it gives up on a particularly useful, and epistemically beneficial function of our ordinary concept. Second, I argue that the conservative notion is not suitable as either a revisionary or a theoretical concept of justification because it in effect changes the subject, and that is not something that our revised or theoretical accounts of normative epistemic concepts should do. Both arguments center around the claim that the conservative picture of evidentialist justification removes – to an unacceptable degree – the possibility of holding one another accountable, as doxastic agents, for important aspects of the belief forming process in which our doxastic agency is involved. This ability to hold one another accountable is a particularly useful and epistemically beneficial function, as well as one of the main social and cognitive purposes of our ordinary concept of justification.

To put us in a position to see what this function consists in, and understand the way in which it is epistemically beneficial, consider the following. We live in an epistemic community. As such, we have an epistemic division of labor. We each have individual faculties for collecting and storing evidence, and forming beliefs. But these are limited. To gain true beliefs about the things we care about, we often rely on the results of other peoples' processes of gathering evidence and forming beliefs. When a friend tells me that \( p \) is true, I believe it not because I have investigated the evidence
for \( p \) independently, but because I trust that she has done her homework, by gathering and weighing her evidence appropriately.

When we acquire beliefs in this way, other people's belief-forming processes act as a sort of stand in for our own. This division of labor reflects an epistemically efficient way of pursuing the truth, but only insofar as the testifiers are using belief-forming processes that we would also want to accept. Thus, we need a way of coordinating our belief-forming processes, both (a) so that other peoples' belief-forming processes are generally acceptable to us as such stand-ins for our own, and (b) so that we can indicate when they are not. Our ordinary concept of justification, and its corresponding terminology, are invaluable in this regard: when we recognize that another person's belief was formed by a process that we would not accept as a stand in for our own, we can express our disapproval by labeling their belief *unjustified*. Applying this label serves to offer an epistemic criticism, indicating that the way the belief was formed or revised is unacceptable on epistemic grounds. And these criticisms influence behavior. Sometimes, we do not have agential control over the relevant aspects of our belief forming processes. But in cases where we do, these expressions have an overall tendency to influence the audience to follow the implicitly endorsed belief-forming rules and practices (or, to refuse to follow those that are not endorsed). When these evaluative practices are used throughout the epistemic community, then, they serve to coordinate belief-forming processes in accordance with (a). In other words, iterated use of the concept of justification in the service of (b) is a means to coordinating belief-forming processes in the service of (a). This coordination is possible because using our ordinary concept of justification lets us criticize one another's belief-forming processes, thus holding one

---

37 I am taking cues here from Sinan Dogramaci's work. For a more detailed explanation, see Dogramaci (2011) and Dogramaci (2015). The way he puts it is that our practice with terms like *rational* functions to extend our collective epistemic reach by enabling each person to serve as an “epistemic surrogate” for any other person.

38 As mentioned in a previous footnote, this includes not only the terms *justified* and *unjustified*, but also terms like *rational, irrational, reasonable, unreasonable*, and perhaps others. These are often used interchangeable to evaluate beliefs in ordinary contexts.
another accountable for those aspects of the belief-forming process in which our agency is involved. 39

In seeking to revise our ordinary concept of justification, it would be foolish to give up on this particular function of that concept, since it is an especially useful one. We need a concept that allows us to coordinate belief-forming processes by holding one another accountable, in order to maintain an efficient division of labor within our epistemic community. It is owing to this arrangement that we have reason to trust that other people's experiences and beliefs can serve as bases for our own, via testimony. So to give up on this feature would be to give up on something that we have epistemic reason to want our concept of justification to do for us. It would be to give up on the possibility of using our concept to push one another to live up to our potential as doxastic agents in those aspects of belief-forming processes in which our agency is involved, and thus achieve an efficient epistemic division of labor.

So, the conservative evidentialist picture of justification is out-of-sync with what we should want from a revision of our ordinary concept, because it gives up on the features of that concept that allow it to function in this epistemically useful way. In §5.2, I will say more about how the conservative picture gives up on these features. The basic idea is that (i) and (ii) point to aspects of the belief-revision process in which the subject's agency is involved, and in which she fails to live up to her potential as a doxastic agent, such that we should not accept her belief-revision process as a stand-in for our own. While our ordinary concept of justification allows us to hold the subject accountable for these aspects of the belief-revision process, the conservative notion does not.

Next I argue that the conservative picture is out-of-sync with what we should want from both revisionary and theoretical accounts of justification. I make a case that in giving up on this key function of our ordinary concept of justification, the conservative picture actually changes the subject. And I argue that we should not want our revisionary or theoretical accounts of normative epistemic concepts to do that.

39 I do not mean that we hold one another accountable in any heavy, moral sense. I mean it in the ordinary sense that we implicitly hold one another accountable when we criticize.
When we are investigating concepts in the revisionary or theoretical modes that I described at
the beginning of this section, there are limits on how revisionary we can be before we simply end up
changing the subject. Put simply: if a revised or theoretical concept departs too far from the original,
then it is no longer the same concept. Why? Without broaching the debate(s) over the structure and
ontology of concepts, we can point out that if a revised or theoretical concept is unable to serve the
main cognitive and social purposes of the original (or at least to serve purposes that are continuous with
them), then there is not enough to ground the revised or theoretical concept as being the same one as
the original. Given that, one can imagine how the argument will go. I will argue that the conservative
evidentialist concept of justification is unable to serve the main cognitive and social purposes of the
original concept. So, the conservative picture of evidentialist justification on which Elena's belief
counts as being proportioned to her evidence does not successfully show that such polarized beliefs are
in fact justified; it simply changes the subject.

The reader may worry that filling in this argument would take quite a lot of doing. Doesn't this
form of argument presuppose a certain picture of the ontology of concepts, on which they have
essential functions that cannot be eschewed? I do not intend to commit to anything this strong. Instead,
I mean to invoke a picture of theoretical philosophy on which an important aim of our theorizing about
concepts is to maintain firm contact with the phenomena to which everyday uses of our words refer.
Here I draw inspiration from Nancy Bauer, who points out that the danger that our use of philosophical
jargon courts is that, in using words that people use all the time outside of philosophy (like
justification), we tend to assume that our work will be relevant to or about the phenomena to which
everyday uses of those words refer. As a result, she says, “We do not feel a standing obligation to
measure the distance between the range of everyday meanings of these words and the meanings we

40 I am borrowing this thought about the importance of a concept's being able to serve main cognitive and social purposes
of the original from Mark Richard, though I am not sure whether he would sign onto the principle I suggest in this form.
philosophers impose on them.” This worry is that in failing to measure this distance, we assume we are making discoveries about real-world phenomena, when in fact we have more or less changed the subject. I myself think that regardless of whether we are collectively deceived about what we are doing in the way that she describes, changing the subject in this way is undesirable. This may be clearer in the case where we are revising ordinary concepts, but I think it holds true for the case where we are developing theoretical concepts, too. In epistemology, we think about the normative import of phenomena that are central in determining how we live our lives. If in doing so, we fail to make contact with how these normative aspects contribute to that determination, then as Bauer puts it, we philosophers fail to seize an opportunity to be relevant in a certain way. Philosophy fails to have the kind of human meaning that it could otherwise.

Even for those who think there is no problem with changing the subject per se, there is reason to think epistemologists in particular ought to feel the kind of standing obligation that Bauer writes about, given the predominant methodologies of our field. Epistemologists routinely draw on our intuitive judgments about cases in developing accounts of normative epistemic concepts like justification and knowledge. Presumably, the reason for doing this is that these judgments reveal something about the contours of our ordinary concepts. If our work is not in large part about or at least relevant to those concepts, then there is, at the very least, reason to worry about the means we have collectively chosen to serve our ends. It is inconsistent to look to our practices of using normative epistemic concepts in order to figure out the correct account of them, but then turn around and say that the accounts we are giving are simply not answerable to the contours of those practices.

For our purposes, we need not look to a sophisticated theory of concepts to measure the proverbial distance between the ordinary and the revised or theoretical meanings of our terms. We can

---

42 For further explanation and defense of this view, see Bauer's (2015), especially chapter 8.
43 Ibid., p.144.
simply rely on the heuristic that if a concept we develop when we are working in a revisionary or theoretical mode is unable to serve the main cognitive and social purposes of the ordinary one – or at the very least to serve purposes that are continuous with them – then it is safe to say we are no longer talking about the same concept. I have already suggested that the conservative notion of justification is unable to serve an important social and cognitive function of our ordinary concept. Namely, our ordinary concept allows us to coordinate belief-forming processes by holding one another accountable for those aspects of the process in which our agency is involved. Absent a concept that functions this way, it is hard to imagine how we could achieve an efficient epistemic division of labor. Why think in addition that it is one of the main cognitive and social purposes of our concept? For one, it is likely that our need of a concept that could play this role in our practice explains why our ordinary concept of justification is what it is in the first place. Our concept was likely shaped in response to this need; thus, it has the feature of allowing us to hold one another accountable in the way I described. In the next section, I will say more about how the conservative concept of justification leaves insufficient room for holding agents accountable in this way.

5.2 The Purpose of Our Normative Epistemic Concepts

In the previous section, I argued that a revised or theoretical account of justification ought to allow us to hold one another accountable for those aspects of our belief-forming processes in which our agency is involved, because that is one of the main purposes of our ordinary concept, and we have reason to want our revised and theoretical notions to be answerable to our ordinary one. In this section, I argue that the importance of having a concept of justification that allows us to hold agents epistemically accountable for aspects of the belief forming process in which their agency is involved goes even deeper. In making this argument, I will also shed further light on how the conservative picture leaves insufficient room for holding agents accountable in this way.
The deeper reason to have a concept of justification that lets us hold doxastic agents accountable in the way I described also stems from the purpose of our concept, in a sense. Stepping back, we can ask about the purpose of a concept in terms of what the sense in having such a concept is, rather than simply in terms of what its useful functions are. What should we want a concept like this for in the first place? At this broader level, the purpose of normative epistemic concepts like justification is to evaluate the beliefs of human doxastic agents – not in terms of how well they promote some external end – but as such. It is natural to think that such concepts should take account of the ways that our human doxastic agency is sometimes deeply involved in the process of forming those beliefs. But in what follows, I will explain how the conservative picture of evidentialist justification leaves precious little room for this. Thus, there seems to be a mismatch between the descriptive facts about how our doxastic agency is involved in belief formation, and the evaluative scheme that this breed of evidentialist offers. To speak metaphorically, the problem with the conservative picture is that it removes the doxastic agent from the evaluative picture to a large degree.

How does the conservative picture remove the agent? Evidentialism about justification starts with the idea that a subject's belief is justified iff it is proportioned to her total accessible evidence. In spelling out what this means, it becomes clear that for the conservative, the evaluative focus is rather exclusively, and seemingly arbitrarily, on this end state of a belief's being proportioned to the evidence. There is no room for taking stock of the ways in which the subject's agency is active in the process, both of doing the proportioning, and of constituting the evidence itself. The evidentialist's evaluative scheme thus seems more appropriate to a descriptive picture on which the doxastic agent is relatively inert, both with respect to the process of proportioning her beliefs to the evidence, and with respect to the evidence itself.

In what follows, I argue that the doxastic agent is not so inert. I make a case that because of this, she has certain kinds of responsibility, both for her evidence, and for the process of proportioning her
belief to it, which ought to be reflected in our concept of justification. Finally, I will suggest that there is in principle room within the evidentialist framework to develop a more liberal picture that takes account of all this in spelling out the normative standard of what it means to proportion one's belief to her evidence. But the concept that results is not one on which polarized beliefs like Elena's will count as justified.

Particularly in complex and controversial cases of the sort where belief polarization is most likely to occur, the doxastic agent has the capacity to be actively involved both in the process of proportioning her beliefs to her evidence, and in the process of constituting the evidence itself. Take the first point first. Earlier in our discussion, I made the point that proportioning one's belief to the accessible evidence can be tricky, such that it requires active cognitive effort on the part of the agent. This is particularly true in complex and controversial cases, where the evidence usually does not come in a neat and tidy package such that the subject can simply observe it and adjust her belief accordingly. To put it more colorfully: it is not as if an oracle reveals to the epistemic agent that there is a 94% chance that \( p \) is true, and all that the doxastic agent has to do in order to proportion her belief to the evidence is adopt a credence of .94. Rather, we have seen that proportioning one's belief to the evidence can be an effortful and active cognitive process. For instance, in our discussion of access, we saw that sometimes evidence is accessible to an agent, though accessing it requires active cognitive effort. In Elena's case, I argued that actively reflecting on how the various pieces of her evidence bear on one another puts her in a position to access the defeat relation between U1 and E2.

A second way in which the subject's doxastic agency is actively involved in the process of proportioning her belief to the evidence is bound up with the fact that that process is, in part, a process of constituting that evidence. Again, particularly in the complex and controversial cases where belief polarization is likely to occur, the type of relevant information that one finds out in the world often does not wear the facts about what it supports on its sleeve. For instance, when Elena is given
information about the congenial and uncongenial studies, she has to do some work to figure out what that information really amounts to vis-à-vis her belief. In Kelly's terms, this work enacts the transition from what he calls narrow to broad evidence. Evidence in the narrow sense consists of relevant information about the world – things that it would be natural to call data. Evidence in the broad sense includes everything of which one is aware that makes a difference to what she is justified in believing. Evidence in the broad sense includes evidence in the narrow sense, but Kelly points out that it also includes things like the space of alternative hypotheses of which one is aware. The information that Elena is given about the congenial and uncongenial studies is narrow evidence. Part of what the process of proportioning her beliefs to the evidence consists in is processing that information, making judgments about how credible it is, and thinking about how much weight it merits in light of any other evidence that she already has. This includes coming up with alternative hypotheses that could account for the data she is presented with, among other things. In doing all of this, the subject's agency is actively involved not only in the process of proportioning her belief to the evidence, but in constituting the broad evidence to which she proportions her belief. To that extent, then, the process of proportioning one's belief to the evidence and of constituting it are one in the same. The subject's doxastic agency is not inert with respect to the proportioning process, or with respect to the evidence itself.

On the conservative picture, these descriptive facts about how the subject's agency is deeply involved in the process of constituting evidence and proportioning her belief to it are not relevant to the epistemic evaluation of those beliefs. All that matters is the final state, and the conservative's rubric for assessing that state bears very little reflection of the agent's role in producing it. There is, for instance, no room for assessing how well Elena did in moving from narrow to wide evidence, or in reflecting on

44 For Kelly, the purpose of making this distinction is to argue that if two subjects have the same narrow evidence, they might nonetheless differ in what they are justified in believing, because they have different evidence in the broad sense Kelly (2008a).
how the different pieces of her evidence bear on one another. She is only held accountable for proportioning her belief to those pieces of evidence that are easily and unproblematically available via introspection or reflection. If the purpose of our normative epistemic concepts is to evaluate the beliefs of beings whose agency is actively involved both in constituting their evidence and proportioning their beliefs to it, then this evidentialist rubric seems greatly impoverished. Indeed, it would seem better suited to a world in which the role of epistemic agency in belief formation is limited to adopting the credence that the oracle recommends. But again, that is rarely how things work in the complex and controversial cases where belief polarization happens.45

So what should our concept of justification be like? For the purpose of evaluating the beliefs of beings whose agency is actively involved in constituting the evidence and proportioning one's belief to it, a more fitting model would leave room to evaluate the agent's performance in these aspects of the belief forming process. Indeed, it should not be controversial to say that subjects do better epistemically when they live up to their potential as doxastic agents in these aspects of the belief forming process. This is why, as I argued in the previous section, we have reason to push one another to live up to our potential as doxastic agents in these ways, and to hold one another epistemically accountable when we fall short. This is not merely a practical matter, as Kelly seems to think, but an epistemic one. So our normative epistemic concepts ought to leave room to account for the fact that we can do epistemically better or worse as a result of these exercises of our cognitive agency.

I have argued with respect to the Elena case, that (i) and (ii) represent an important part of what determines our judgment that her belief is unjustified. She could have done better epistemically by processing her narrow evidence more scrupulously, and by being more reflective in taking stock of how

---

45 The conservative concept of justification may be better suited to evaluating less complex cases, like perceptual belief. There, the descriptive facts are more closely aligned with the oracle model of proportioning one's belief to the evidence. Generally, considering one's perceptual evidence does not require intentional cognitive action on the part of the epistemic agent. So it may be that the conservative evidentialist had these cases in mind in constructing her theory of justification. But I would argue that the model is not appropriately extended to these complex and controversial cases.
the different pieces of her evidence bear on one another, and of what they support as a whole. Indeed, this explains why we are compelled to call Elena's belief but not Elena2's unjustified: Elena2's agency is not complicit in the bad process by which her belief was revised. The conservative picture does not leave room to account for this. But there is, in principle, room to account for it within a theory of justification that is broadly evidentialist. All it would require is that there be a normative standard of what we expect from the agent embedded in our understanding of what it means to successfully proportion one's belief to her total evidence. This standard might give up on the restriction that a belief cannot be criticized for (i), (ii), or both. Giving up on (i) would allow evidentialism to hold the subject to a standard of what it means to succeed at the part of proportioning one's belief to the evidence that involves considering one's narrow evidence in order generate wide evidence. This would mean that the agent is normatively responsible not just for the evidence that she in fact has at the moment when she proportions her belief to it, but also for whatever evidence she ought to have as a result of meeting this standard. Giving up on (ii) would allow evidentialism to hold the subject to a standard of what it means to put adequate effort into the part of proportioning one's belief to the evidence that involves turning inward and taking stock of one's evidence, how its different pieces might bear on each other, and what it amounts to as a whole. I argued that holding the subject to such a standard should entail holding her responsible for the defeat relationship between U1 and E2. More broadly, it would mean that accessible evidence is a normative concept, whose specific content should be determined in part by what the relevant type of creature is cognitively capable of. For instance, competent adult human doxastic agents are fitting subjects for a higher standard of what it means to reflectively take stock of one's evidence than young children and nonhuman animals are, because the latter kinds of subjects lack

46 The specific question of whether the evidence that Elena would be normatively responsible for on this kind of standard includes the evidence that she would have gotten from scrutinizing the congenial study is beyond the scope of our discussion here. But I do not think it would be crazy to suppose that her belief loses some measure of justification because she lacks this evidence. After all, Elena does not have any positive reason to think that the congenial study will be free of flaws when she chooses not to scrutinize it.
the type of cognitive agency that I have been describing.

More generally, understanding the evidentialist definition of justification in the suggested way means that the concept of evidence in play is a normative one. A person's belief is justified iff it is proportioned to her total evidence, where that is the evidence she ought to have as a result of meeting a certain standard of exercising her doxastic agency in the belief forming process. Because the subject's agency is involved in constituting her broad evidence, and actively reflecting on what its overall weight and balance supports, she bears a normative epistemic responsibility for the evidence that she ought to have as a result of doing these things well.

In a sense, the conservative picture of justification already has a standard of what we should expect from the doxastic agent embedded in the definition of what it means to successfully proportion one's belief to her total evidence. It is just a lower one. Presumably, for instance, Kelly would consider Elena's polarized belief unjustified if she found the fatal flaw in the uncongenial study, remained aware that it had such a flaw in the moment of revising her belief, but failed to treat it as a defeater of the support that E1 would otherwise lend to Not-Deterrent. This would count as a failure to successfully proportion her belief to her total evidence, since although she recognized that the uncongenial study was fatally flawed, she failed to recognize the way in which this bears on E1, and to proportion her belief to her total evidence in light of that connection. Similarly, I imagine he would consider Elena's belief unjustified if in moving from narrow to broad evidence, she found the fatal flaw in the congenial study, but then willed herself to forget that she had seen it, so that her broad evidence base continued to reflect that the congenial study's results support Deterrent.47

In §4.1, I said that an evidentialist like Kelly would not be impressed by an argument that

47 It is unclear just how low the standard is on the conservative picture, but there is reason to think that it is pretty low. After all, Kelly does not take issue with Elena's judgment that the congenial study does not contain any serious flaws, even though she has been presented with information that direction challenges this, in the form of criticisms of the congenial study. This might mean that an error in the opposite direction would also be deemed unproblematic – for instance, if the uncongenial study contained only a minor flaw, and Elena judged it to be a fatal one. This seems like the sort of thing that could problematically lead people to dismiss entire bodies of research in the wild.
epistemic irresponsibility during the process of belief formation can undermine the epistemic status of resulting beliefs. But in a way, what I have done here is to suggest that Kelly can make his picture of evidentialist justification more tenable by writing a concept like epistemic responsibility back into the definition of justified belief. What we are now in a position to see is that this idea need not be considered altogether foreign to the evidentialist framework. We just saw that the conservative picture already has a certain standard of responsibility written into the definition of what it means to successfully proportion one's belief to her total evidence (albeit a relatively deflated one). To that extent, it does matter, even to the conservative evidentialist, whether the subject exhibits epistemic responsibility in the process of acquiring evidence and proportioning her belief to it. The difference between the conservative's conception and this more liberal one lies in their different understandings of the range of our epistemic responsibility. The more liberal definition claims that our condition and capacities as doxastic agents render us fitting subjects for a broader range of responsibilities. Ultimately, a careful consideration of just how far this range of responsibility reaches is beyond the scope of my discussion here. Earlier, I argued that there is reason to think our ordinary concept of justification already holds us responsible for (i) and (ii), and that we should be wary of giving this up. And in the present section, I have argued that there is reason to think we are fitting subjects for this broader range of epistemic responsibility, given a metaphysical picture of the capacities we have as doxastic agents, and how they are involved in belief formation. I leave the question of exactly how to delimit the range of our responsibility aside, though I think it ought to be based on our best understanding of our capacities as doxastic agents, and the ways in which exercising them allows us to do well for ourselves epistemically.

If what I have said here is right, then the choice for Kelly is between adopting a more liberal definition of evidentialist justification, or conceding that Elena's polarized belief is not justified on evidentialist grounds. Given what I have said about the tenability of the conservative definition, I
advise the former.
1. Introduction

1.1 Theories of Epistemic Justification

At a broad level, the debate over the correct theory of epistemic justification is about what sort of things can do the justificatory work. We can arrange the various views according to how restrictive they are about this, as follows:

- **Externalism**
- **Mentalism**
- **Internalism**
- **Accessibilism**
- **Articulability**

The most restrictive view about what things can do the justificatory work is a theory of justification that I will call *articulability*. On this view, only reasons that a subject can *articulate* count as reasons that can justify her belief. This is not an overwhelmingly popular view in the literature, but Adam Leite's theory of justification seems to be something quite close to this. He is motivated by the thought that, as he puts it, it would be outrageous dogmatism for a person to insist upon the truth of her belief even though she cannot satisfactorily defend it. And he thinks this supports the idea that the epistemic status of a belief that a subject insists upon in this way must be tied to her poor performance in the activity of justifying.

48 More properly, *articulabilism*. But that does not exactly have a nice ring to it.
Ultimately, his view of justification is that for a person to be justified in believing some proposition, she must be able to develop and provide an appropriate and adequate defense of her belief when asked to do so under appropriate conditions.\footnote{By “appropriate conditions”, he simply means, for instance, that it is not the case that the subject is experiencing temporary amnesia, exhaustion, neurosis, repression, &c.} More specifically, “On this view, to be justified is to have a sufficient grasp of how the world is (including both the relevant facts and the relevant relations of evidential support), and sufficient rational capacities, to enable one to recognize good reasons which one possesses for believing as one does, to make them one's reasons for holding the belief, and to respond adequately to objections.”\footnote{Ibid., p.242.}

The next most restrictive view about what sort of things can do justificatory work is accessibilism. The accessibilist holds that in order for a subject to be justified in believing a proposition, she must have actual or potential access to (that is, awareness of) something that justifies it. Weaker versions of accessibilism only require potential access, whereas stronger ones require actual access.\footnote{There are plenty of other fault lines along which we can divide stronger versions of accessibilism from weaker ones. For instance, weaker versions might only require access to the justifiers, whereas stronger ones might require access to the fact that the justifiers are in the basis for one's belief, or even access to the fact that because of this, the belief is justified. I will focus on a particular version of accessibilism, which is spelled out in §2. The present aim is simply to get a sense of how some theories of justification are more restrictive than others about what sort of things can do the justifying.}

Accessibilism is often contrasted with a less restrictive theory of internalist justification, known as mentalism. Mentalism says that justification is determined entirely by states that are internal to the subject's mind. This is different from accessibilism because although the justifiers must be mental states, the subject need not have access to them. Both accessibilism and mentalism are varieties of internalism in the sense that the factors that determine justification are entirely internal to the subject's mind.

The least restrictive view of what sort of things can justify a subject's belief comes from externalism about justification. For the externalist, nonmental factors, even those that are completely
external to the subject, can do this justificatory work. For instance, if a deity of some sort is causing beliefs in a subject, there is no in principle reason why that type of input could not count as justificatory.

1.2 The Received View

My primary focus will not be on the debate between internalism and externalism, but rather on a debate within internalist theories of justification. If we are going to be internalists, what kind of internalists should we be? Which internalist position(s) are defensible?

In the literature on internalist theories of justification, there has been very little discussion of mentalism as compared with accessibilism. And those who have addressed it have often done so in order to say that it is not worth defending. They have given a couple of different reasons for this. Conee and Feldman, who are well known internalists, have suggested that it is likely that philosophers have not often separated mentalism from accessibilism because they have tacitly assumed that their extensions do not differ in any significant way. They explain this in the following way:

“...philosophers...have tacitly assumed...that the special kind of access on which many internalist theories rely can reach only mental items, and perhaps all mental items, or at least all that might be counted as playing a role in justification. The italicized part of the claim is non-trivial. To be fair, Conee and Feldman do not explicitly endorse the tacit assumption they identify here. Nonetheless, they do go on to defend internalism as if it were a single, unified theory, rather than separating mentalism from accessibilism. Part of my task in this paper will be to examine this non-trivial part of their claim, and ultimately, to argue that there is reason to think it is false.

Another reason it has been espoused that mentalism is not worth defending comes from Michael Bergmann, who says that to go mentalist is to give up on the main motivation for being an internalist

---

about justification in the first place. More specifically, he thinks that giving up on the accessibilist's awareness requirement eliminates the clearest and most compelling motivation for endorsing internalism. Although Bergmann does not ultimately subscribe to an internalist theory of justification, he does think that on the face of it, it has a powerful appeal; but, he says, it only has that kind of appeal if by internalism, we mean accessibilism. Thus, he argues that in the end, we should not even think of mentalism as a kind of internalism, since it gives up on the main motivation for endorsing internalism. I will argue that this is incorrect. I will spell out the motivation for internalism that Bergmann has in mind, and argue that ultimately, it does not give us reason to rule out mentalism.

The received view is that mentalism, as distinct from accessibilism, is not worth defending. I will argue against this that mentalism had merit as a version of internalism about epistemic justification. In §2, I spell out the motivation for internalism that Bergmann proposes. Ultimately, I argue that when we think about what it really comes to, we see that it does not rule out mentalism as a tenable theory of justification. In §3 and §4, I offer different types of cases in which our intuitive judgments about justification suggest that having access to our justifiers is not necessary for justification. In §5 and §6, I discuss some implications for our social epistemic practices, and make a suggestion for future inquiry.

2. The Motivation for Internalism

For Bergmann, although internalism about justification does not ultimately succeed, the most compelling motivation for endorsing an internalist view comes from its ability to respond to a certain kind of objection. Indeed, the structure of his ultimate argument against internalism relies on the idea that once we eliminate this, there are no other compelling reasons for endorsing internalism to replace it. He refers to this objection as the Subject's Perspective Objection (SPO), which he formulates as

follows:

SPO: If the subject holding a belief isn't aware of what the belief has going for it, then she isn't aware of how its status is any different from a stray hunch or an arbitrary conviction. From that we may conclude that from her perspective it is an accident that her belief is true. And that implies that it isn't a justified belief.  

55

Ultimately, I will problematize the thought behind this objection. But let us first follow along with Bergmann in order to get a better idea of his reasoning here, and of what he thinks it would take to get around the objection. The starting point on which we can all agree is the idea that a stray hunch or an arbitrary conviction lacks the epistemic status of a justified belief. So, a theory of justification ought to be able to give an account of what it is that makes a justified belief epistemically superior to a stray hunch or arbitrary conviction. The idea behind the SPO is that the only reasonable candidate for what this difference-maker might be is an awareness on the part of the subject of how the status of her belief is different from that of a stray hunch or an arbitrary conviction.

What does it take to have such an awareness? Bergmann starts by outlining the basic requirement, and then goes on to refine it further. At that first stage, he explains that the accessibilist is able to avoid the SPO by imposing an Awareness Requirement on justification:

The Awareness Requirement: S's belief B is justified only if:

(i) There is something, X, that contributes to the justification of B – e.g. evidence for B or a truth-indicator for B or the satisfaction of some necessary condition of B's justification; and,

(ii) S is aware (or potentially aware) of X.  

56

X will be some mental item that contributes to the belief's justification. And the type of awareness that S has of X will be what Bergmann calls strong, as opposed to weak awareness. Strong awareness is an awareness that involves conceiving of X as being in some way relevant to the truth or justification of the belief, whereas weak awareness would not involve conceiving of it that way (weak awareness

56 Ibid., p.9.
might involve conceiving of it in some other way, or it might involve a non-conceptual awareness). So, in order to avoid the $SPO$, the subject must be aware of some mental item that in fact contributes to her belief's justification, and she must conceive of it as relevant to that belief's justification or its truth.

Bergmann goes on to consider various more specific versions of the awareness requirement. His end game is to show that if the awareness required by internalism is strong awareness, then internalism has a vicious regress problem leading to radical skepticism; and, if the awareness required by internalism is weak awareness, then internalism is vulnerable to the $SPO$. So, he considers various more specific versions of the awareness requirement that involve strong awareness, and argues in each case that the requirement leads to a vicious regress. But our only concern for present purposes is to understand what the awareness requirement would need to be like in order to avoid the $SPO$.

The purpose of the awareness requirement is to guarantee that the believer is aware of how the epistemic status of her belief is different from that of a stray hunch or arbitrary conviction. As stated, it is too weak to guarantee this. It needs to be modified in a couple of ways.

Take the second condition of the requirement first. It says that the subject is aware or potentially aware of some justification-contributor. Presumably, Bergmann includes this disjunct in his formulation of the requirement because it is usually included in standard formulations of accessibilism. But I am not convinced that it is useful for the purpose of avoiding the $SPO$. If the function of being aware of a justification-contributor is to guarantee that when a person's belief is justified, she is aware of how its epistemic status is different from that of a stray hunch or an arbitrary conviction, then merely having the potential to become aware of that does not seem like enough. Potential awareness only implies that the subject could become aware of this, were some non-actual conditions to obtain. To secure the guarantee that allows us to avoid the $SPO$, we need actual access.

In his book, Bergmann raises the further question of whether the strong awareness that is required in condition (ii) is a doxastic or a nondoxastic strong awareness. Doxastic strong awareness is
strong awareness that involves the belief that the object of awareness is in some way relevant to the truth or justification of the relevant belief. Nondoxastic strong awareness is just strong awareness that is not doxastic, but still involves the kind of conceiving that is characteristic of strong awareness.\textsuperscript{57} Bergmann draws here on Paul Moser's idea that there is a distinction between doxastic (or propositional) awareness of X, and mere conceptual awareness of X.\textsuperscript{58} Propositional awareness of X involves a judgment predicating something of X, whereas mere conceptual awareness of X involves categorizing X according to some classificatory scheme, without any judgment being made. Bergmann ultimately argues that, for his purposes, it does not matter whether the internalist imposes a doxastic or a nondoxastic strong awareness requirement on justification, since both kinds of strong awareness lead to vicious regress, and thus to radical skepticism. For my purposes, I will adopt the weaker, nondoxastic strong awareness requirement, in order to formulate the weakest version of the requirement that can avoid the \textit{SPO} by guaranteeing that when a subject's belief is justified, she is aware of how its epistemic status is different from that of a stray hunch. This awareness does not require a belief about X. It simply requires that the subject conceive of X as something that contributes to the belief's justification.

Returning to the first condition of the awareness requirement: note that Bergmann lists a few different possibilities for X, as things that could contribute to B's justification. For reference, they are: “Evidence for B or a truth-indicator for B or the satisfaction of some necessary condition of B's justification.” The requirement that the subject is aware of one of these things, even as something that contributes to B's justification, is not very demanding. Again, if the purpose of the awareness requirement is to guarantee that when a subject's belief is justified she is aware of how its epistemic

\textsuperscript{57} Ibid., p.14-17.  
status is different from that of a stray hunch or an arbitrary conviction, then this requirement is not strong enough. Take the first two possibilities for X first. Being aware of evidence or a truth-indicator for B is not sufficient for being aware of how B is different from a stray hunch or an arbitrary conviction, unless we define a stray hunch or an arbitrary conviction as something for which one does not have \textit{any} evidence.\footnote{For present purposes, I am treating \textit{evidence} and \textit{truth-indicator} as synonymous. They are both justification-contributors, which seems to be what is relevant here.} And that does not seem right. A belief could still count as an arbitrary conviction if the subject has some evidence, so long as it clearly falls short of what would be needed for justification. Think of the old man who gets a feeling in his bones when it is going to rain. Suppose that as he gets to be an even older man, he finds increasingly that he gets that same feeling in his bones for other reasons, too. He gets it anytime the weather changes, he gets it when he has been walking around too much, he gets it when he has been walking around too little, he gets it when he has not been eating enough protein, and sometimes he gets it for no reason at all. Suppose that one day he gets the feeling, and for no additional reason, he has a hunch that \textit{this} time, it indicates rain. Further suppose that he conceives of his feeling as justification-contributing evidence for his belief that it will rain. He is not wrong to do so. The feeling is \textit{some} evidence for the belief. But it is not enough to distinguish the belief from the realm of stray hunches and arbitrary convictions.

Bergmann's third possibility for X is that the subject is aware of the satisfaction of some necessary condition of B's justification, as something that contributes to its justification. But this also falls short of guaranteeing that the subject is aware of how the status of her belief is different from that of a stray hunch or an arbitrary conviction. Indeed, necessary conditions come cheap. For instance, one necessary condition of a subject having a justified belief is that she has a belief. And in the absence of defeaters, the subject would be right to think of the satisfaction of this condition as a justification-contributing evidence: our belief-forming mechanisms are designed to produce justified beliefs, so the fact that one finds herself with a belief is \textit{some} evidence that it is justified. But it does not follow that...
the subject is aware of how the status of her belief is different from that of a stray hunch or an arbitrary conviction. To secure the guarantee that allows us to avoid the SPO, we need to make the awareness requirement stronger, by appending an additional requirement on how the subject must conceive of X. Namely, she must not only conceive of it as being something that contributes to B's justification, but also as something that could not easily attach to a stray hunch or an arbitrary conviction.

Taking all of this into consideration, I suggest that a theory of justification can avoid the SPO by adopting the following Modified Awareness Requirement on justification:

**Modified Awareness Requirement:** S's belief B is justified only if,

(i) There is something, X, that contributes to the justification of B – e.g. evidence for B or a truth-indicator of B or the satisfaction of some necessary condition of B's justification; and,

(ii) S is aware of X in such a way that S justifiedly applies to X the concepts of being something that contributes to B's justification, and being something that could not easily attach to a stray hunch or an arbitrary conviction.

The only differences between this formulation of the modified awareness requirement and Bergmann's original requirement are in condition (ii): the modified version requires that S have actual, strong awareness of X, and that she justifiedly applies to X the two concepts listed. Bergmann explains that it is important that S justifiedly apply these concepts to X because, as he puts it, “If you think insane or irrational concept application is sufficient, why think concept application is even necessary?” The question of what exactly makes the application of a concept justified is beyond our scope here, but I will work with a liberal and commonsense understanding of what it takes for the application of a concept to be reasonable.

In what follows, I introduce a couple different types of cases in which the modified awareness requirement is not fulfilled, and yet intuitively, the subject's belief is justified. I argue that this gives us reason to re-think the SPO, by asking whether it really constitutes an objection at all. It also gives us

---

reason to think that the Conee and Feldman point is wrong. That is, it looks like we do not have access to all mental items that might be counted as playing a role in justification, so mentalism and accessibilism are not extensionally equivalent. Ultimately, I argue that this puts the burden of reasonable argumentation back on the accessibilist, to provide a principled reason that a belief cannot count as justified unless the subject is aware of how its epistemic status is different from that of a stray hunch or an arbitrary conviction.

3. Thin Slicing Cases

The first type of case in which I think the modified awareness requirement is not met, but intuitively the subject's belief is justified, comes from a psychological phenomenon known as thin-slicing. The term was coined by psychologists to refer to a phenomenon where in certain situations, people are able to make perceptual judgments quickly, reliably, and seemingly with very little consciously-available information, but in a way that makes those judgments reliably true. Often times, the perceptual bases for such judgments are rooted in extensive patterns of exposure, but are unavailable to consciousness. So beliefs based on thin-slicing are unlike paradigmatic cases of perceptual belief in that much of the information on which the beliefs are based is not consciously available to the agent. The term thin-slicing comes from the idea that these judgments are based on a 'thin slice' of experience, as opposed to a long and careful process of deliberation about one's evidence.

What follows is a paradigmatic example of thin-slicing. The case is relayed by Jonah Lehrer in his book on decision making. Before I paraphrase the story, a caveat: as the reader may be aware, Lehrer's work became the center of a scandal a few years ago, in which it was uncovered that certain content that he had presented as factual was exaggerated, or altogether fabricated. Unfortunately, it turns out that he did distort the details of this story to some extent. This distortion will not have any

61 Lehrer (2010).
negative impact on my argument, but I will relay both versions of the case so that the reader may judge for herself. First, here is my paraphrasing of the way that Lehrer tells the story.

Lehrer's version of the Riley case:

The story takes place in February 1991, during the Gulf War, in the context of Operation Dessert Storm. Lieutenant Commander Michael Riley had been charged with the task of monitoring the radar responsible for protecting a fleet of Coalition battleships off the Kuwaiti coast. Early one morning, he was faced with a unique challenge. He had been on duty all night, watching dozens of radar blips on the display, which were friendly fighter jets flying through the airspace surrounding the naval convoy. At 5:00 in the morning, a blip appeared on the screen that immediately filled him with a strong sense of fear. His heart started to race, and his hands became clammy. But Riley did not understand why, since for all he could tell, the blip he was staring at seemed to look and move exactly like the friendly fighter jet blips that he had been watching all night. Baffled by his own reaction, Riley continued to stare at the blip for another minute, but he remained unable to discern any way in which it looked different from all the others.

Soon, Riley was out of time. The blip was headed straight for the USS Missouri. He knew that if it was a missile, the ship would be sunk unless he acted. So he issued the order to fire, and shoot the blip down. Moments later, the captain burst into the room and asked Riley how he could be sure he was firing at an Iraqi missile and not an American fighter jet. Riley said he just knew.

As it turns out, Riley was right. But from his own perspective, he had no idea what his belief that the blip was dangerous was based on. Later, cognitive psychologists discovered that there was a perceptible difference between the friendly blips and the dangerous one. Because missiles travel at a lower altitude than planes do, the signal of the missile was initially masked by ground interference. This meant that it did not appear until the third radar sweep, whereas a plane would have appeared after the first one.
The corrected version of the Riley case:

As it turns out, Lehrer got the Riley story from Gary Klein, who tells what is presumably the true version in his own book on decision making.\(^62\) Klein's version reveals that Riley *did* get more information about the blip that frightened him before making the decision to fire at it. Ordinarily, it takes 30 seconds for the radar system to relay altitude information, but in this case, the weapons director's failure to type the track number correctly, which meant that Riley did not get altitude information until 44 seconds after the relevant blip appeared on the screen and filled him with fear. At 44 seconds, he learned that the target was flying at 1,000 feet. This confirmed his intuition that the blip was dangerous, and only then did he issue the order to fire.

The difference between the true story and Lehrer's distorted version is not particularly important for my argument. Our concern is with whether Riley's belief that the blip was dangerous is justified, not whether his decision to fire at it was rational. We can ask this question with respect to the time-slice version of Riley as he was 43 seconds into the incident, before he got the additional altitude information. I will suppose that he also had the belief that the blip was dangerous at that point, but had not yet decided whether to act on it.

3.1 The Justification Question

Intuitively, Riley’s belief that the errant blip was dangerous is justified. I will argue that while mentalism can account for this, the type of accessibilism that we are considering cannot. But let me first say something about what I think the intuitive judgment that Riley's belief is dangerous is responding to.

The going theory about why thin-sliced judgments are accurate is that they have a perceptual

\(^{62}\) Klein (1999).
See also: <https://reason.com/archives/2012/09/20/welcome-to-the-golden-age-of-fact-checki/1>
basis that is rooted in extensive patterns of exposure, but is unavailable to conscious awareness. As applied to our case, this would mean that Riley is drawing upon a kind of perceptual *expertise*, even though he is not aware of what it is that he is responding to. In other words, Riley's belief that the blip is dangerous is based on mental events of unconscious perception that are informed by expertise about what these blips normally look like. More specifically, the idea is that he has some unconscious mental states that represent the fact that this particular blip looks different from what friendly plane blips normally look like. And this is what I believe underlies the intuitive judgment that Riley's belief is justified: but for the fact that these mental states are unconscious, this looks like a paradigm case of a well-formed belief. The subject is responding to or picking up on a fact about the world, and his belief is caused by and based on that response in a way that is non-accidental and non-deviant. At that level of description, this is a paradigm case of an epistemically well-formed belief. The only difference is that in the Riley case, his response is unconscious. This makes the case a particularly useful one, since it allows us to isolate this one factor, and observe that it does not seem to be a difference-maker for our intuitive judgments about justification.

While mentalism has the resources to account for the idea that Riley's belief is justified, the type of accessibilism that we are considering here does not. This type of accessibilism says that a belief is justified only if it meets the modified awareness requirement. As applied to our case, this would mean that Riley's belief that the blip is dangerous is justified only if: (i) There is something, X, that contributes to the justification of his belief B – *e.g.*, evidence for B or a truth-indicator of B or the satisfaction of a necessary condition of B's justification; and, (ii) Riley is aware of X in such a way that

---

63 There are a few different options for what exactly these unconscious mental states are. It could be that there is an unconscious perception whose content includes the idea that something is amiss with this blip. Or, it could be that there is an unconscious perception whose content does not include this idea, and a further unconscious inference based on an unconscious comparison of this representation of the missile blip with a stored representation of what plane blips normally look like. On the latter option, the unconscious perception of the missile blip might only include an indeterminate or underspecified set of features, or, it might represent the fact that the missile blip appeared after 3 sweeps.
he justifiably applies to X the concepts of *being something that contributes to B’s justification* and *being something that could not easily attach to a stray hunch or an arbitrary conviction*.

In Riley's case, there are not many candidates for X, because there are not many things he is aware of that might count as justification-contributors. Things that might plausibly count are: his fear reaction to the blip, his inclination to believe that it is dangerous rather than friendly, and the fact that he has not reacted this way to other blips. Do any of these meet condition (ii)? That is, are any of them things to which he justifiably applies the two concepts? In each case, it is debatable whether Riley justifiably applies the first concept, of *being something that contributes to B’s justification*. The fact that Riley is so hesitant to act on his belief – especially in the true version of the story – is evidence that he does not conceive of any of these things as being *sufficient* for justification. But he may well think of them as *contributors*. And indeed, this does not seem unreasonable (on a commonsense and liberal understanding of what it takes to justifiably apply a concept, we need not set a particularly high bar on what it takes to justifiably conceive of something as a justification-contributor). But are any of these things to which he *also* justifiably applies the concept of *being something that could not easily attach to a stray hunch or an arbitrary conviction*? It does not seem so. Indeed, it does not seem that Riley conceives of any of his justification-contributors this way at all, regardless of the question about justification. This would explain why he continues to stare at the blip for 44 seconds without acting. Because he does not conceive of the justification-contributors to which he has access as things that could not easily attach to a stray hunch or an arbitrary conviction, he is not aware from his own perspective of how the epistemic status of his belief is any better. And so he is reluctant to act on the belief. Instead, he continues to stare at the screen, searching for some further X that he can justifiably conceive of as a justification-contributor that could not attach to a stray hunch or an arbitrary conviction, so that he can be aware of how the belief he acts on has a different status. Until he finds one, his belief does not fulfill condition (ii) of the modified access requirement and thus does not avoid
the SPO. So, the breed of accessibilism whose aim is to avoid the SPO does not have the resources to explain the intuitive judgment that Riley's belief is justified.

Mentalism, on the other hand, can explain the idea that Riley's belief is justified, since the mentalist can count Riley's unconscious, inaccessible mental states as securing justification. Again, the thought is that Riley has some unconscious mental states that represent the fact that this particular blip looks different from what friendly plane blips normally look like. These inaccessible states contribute to causing the accessible ones, like his fear reaction and inclination to believe that the blip is dangerous. And for the mentalist, Riley's belief can derive its justification from these inaccessible informational states. It does not follow from their being inaccessible that they are not sufficient for justification.

So, mentalism is able to capture the intuitive idea that Riley's belief is justified, whereas the variety of accessibilism that has a modified awareness requirement on justification is not. And the accessibilist analysis of the case conflicts with our intuitive judgment because that judgment is responding to the very feature of the situation that accessibilism covers over; namely, the importance of Riley's unconscious mental states. Our intuitive judgment picks up on the way in which these states make a justificatory difference. For the mentalist, it need not be that the subject is aware of how the status of her belief is different from that of a stray hunch or an arbitrary conviction. It is simply that its status must in fact be different. So it is not a problem that the difference-makers are unconscious and inaccessible.

3.1.1 Externalism

Though my primary focus here is on a debate within internalist theories of justification, this point about the justificatory role of our unconscious mental states also helps to explain why an externalist theory of justification does not give us the right result. Consider a different version of the
story, starring Clairvoyant Riley. Clairvoyant Riley's situation is exactly the same as actual Riley's, except that his belief that the blip is dangerous is not a result of thin-slicing, but rather of clairvoyance. Clairvoyant Riley observes a blip on the radar screen, and experiences the same feeling of fear, and the same inclination to believe it is dangerous as actual Riley does. But in this case, it turns out that there is no difference in the blip. The blip that Clairvoyant Riley responds to in this way looks identical to a friendly blip. But just as in actual Riley's case, the blip turns out to be a missile. It only looks identical to a friendly blip because of a glitch in the system.

So, Clairvoyant Riley's conscious experience is the same as actual Riley's: he stares at the blip for the same amount of time, unable to discover any way in which it looks different from the plane blips (though in this case, of course, there is nothing to discover). As a result of his clairvoyance, Clairvoyant Riley forms the belief that the blip is dangerous. As it happens, his clairvoyance is perfectly reliable, but he does not know that he is clairvoyant.

In this case, the mentalist does not have the resources to say that Clairvoyant Riley's belief is justified, since there are no mental states to act as justifiers. But on a certain popular theory of externalist justification – namely, *process reliabilism*, Clairvoyant Riley's belief is justified. The simplest version of process reliabilism says that a subject's belief is justified iff it is formed by a reliable process. Since Clairvoyant Riley's clairvoyance is perfectly reliable, process reliabilism would say that his belief is justified.64

Intuitively, Clairvoyant Riley's belief that the blip is dangerous is not justified. I suspect this is because it is not different in the right way from a stray hunch or an arbitrary conviction. The presence of Riley's mental states was a key difference-maker in this regard. So externalism (or at least simple

---

64 Since the debate between internalists and externalists is not my primary topic, I will not concern myself here with testing the Riley cases against more complex versions of reliabilism. Jack Lyons, for instance, has developed a more sophisticated reliabilist theory that would not give the result that Clairvoyant Riley's belief is justified. But with any such theory, we should ask ourselves whether the locus of justificatory power comes from the reliability itself, or from some other source. See Lyons (2009) and (2011).
process reliabilism) does not give us the right result in this case either.

4. Hermeneutic Injustice Cases

A second type of case in which the subject's belief is intuitively justified even though the modified awareness requirement is not met comes from a phenomenon that philosopher Miranda Fricker has termed *hermeneutic injustice*.\(^{65}\) The term refers to a type of situation that a person can be in. I will suggest that people in situations of hermeneutic injustice have justification for certain beliefs that they form about their experiences, and that while the mentalist can account for this, the type of accessibilist we are considering cannot.

In her book on epistemic injustice, Fricker defines hermeneutic injustice as “The injustice of having some significant area of one's social experience obscured from collective understanding owing to a structural identity prejudice in the collective hermeneutical resource.”\(^{66}\) In other words, hermeneutic injustice is a phenomenon that occurs when certain marginalized agents – that is, people who are members of groups that experience a relative lack of social power – struggle to make sense of experiences that they have had, and to communicate about them owing to the fact that, collectively, we lack the concepts required to properly describe and understand those experiences. And importantly, these are cases in which we lack the relevant concepts *because* of the way these groups are socially marginalized. Our collective conceptual (*i.e.*, hermeneutical) resources are heavily influenced by the experiences of groups that have social power, and relatively uninfluenced by those of groups that lack it. As a result, the conceptual tools that we have for understanding our experiences and communicating them to one another disproportionately represent the experiences of the more socially powerful groups.

An example will help to make the concept of hermeneutic injustice more intuitive. Fricker's canonical example has to do with the concept of *sexual harassment*, which became part of our cultural

\(^{65}\) Fricker (2009).
\(^{66}\) Ibid.
currency during the 1970's. The phenomenon of sexual harassment, of course, had been around much longer. But as Fricker explains, before the 1970's, there was no concept for it. Because of this, the phenomenon was invisible in a certain way, even to the very people who were experiencing it. When instances of sexual harassment occurred, they were often normalized as something for which there was an available concept, like flirting or flattery. More generally, the behavior was understood not as a kind of harassment, but as something that simply comes with the territory of being a woman in the workplace. As Gloria Steinem famously put it in a 2011 interview: “...You know, when we started, there was not even a term for sexual harassment – it was just called life.” As a result, the fact that the behavior was part of a larger social pattern, and the ways in which that pattern was socially and morally problematic, were not yet part of a concept in terms of which the people experiencing sexual harassment would normally understand that experience. So, people in this situation of hermeneutic injustice had trouble understanding their own experiences and explaining them to others. In what follows, I describe a particular example of hermeneutic injustice. I argue that the subject intuitively, the subject in this example is justified in believing certain propositions about her experience, and while the mentalist is in a position to explain this, the accessibilist is not.

Carmita Wood's Sexual Harassment Case:

Fricker explains in her book that the term sexual harassment first appeared in the 1960's, in relation to the case of a woman named Carmita Wood. Wood worked in an administrative position in Cornell University's department of nuclear physics, and was being repeatedly sexually harassed by her boss. But as we said, during the 1960's, our collective hermeneutical resource did not yet include a concept that described this behavior. In some sense, Wood was clearly aware of what was happening to her. She was very bothered by it, so much so that she eventually quit the job, even though she was in a

67 See <http://www.interviewmagazine.com/culture/gloria-steinem/>
68 Fricker (2009), p.150.
position of being the sole source of financial support for her two children. But in another sense she did not fully understand her experience, since we collectively lacked the conceptual resources to explain the behavior in terms of how it was morally and socially problematic. After she quit the job, Wood applied for unemployment insurance. When the claims investigator asked her why she had quit her job after 8 years, she was at a loss for how to explain it. Eventually, she said that she had quit for 'personal reasons', and her claim was denied.

There is a whole host of things that Wood probably believed about her experience, even though she lacked the concept that would figure in content of the belief “I am being sexually harassed”. Suppose that she formed the belief, about her boss, “this person has wronged me”. Does she have justification for that belief? Intuitively, she does. In what follows, I make the case that while the mentalist has the resources to explain this, the accessibilist does not.

Of course, whether Wood has mentalist or accessibilist justification for the belief is going to depend on how we fill in the details of what it is like for her, as the person undergoing the experience of harassment. There are many ways that it might be, and presumably victims of sexual harassment have a variety of different experiences. For our purposes, I will stipulate what it is like in a way that allows us to make the distinction between mentalist and accessibilist justification. Even if this is not in fact what it was like for the actual Carmita Wood, it describes a way that the experience could be for some person who experiences sexual harassment in a situation of hermeneutic injustice.

Consider the first instance in which Wood is sexually harassed by her boss. Suppose she walks away from the experience with an intense but inchoate feeling of discomfort, but she does not know why she feels that way, or even how to describe what it feels like. Try as she might, she cannot think of anything about the experience she just had that should have made her feel so strangely – it was all just normal office behavior, right? Still, she cannot help feeling that she has been wronged somehow. But again, try as she might, she cannot understand why, since all that she has conscious access to is the fact
that she is uncomfortable, and the experience of what that is like. She finds that she cannot rule out the possibility that her feeling of discomfort was caused by something unrelated, and her feeling that she has been wronged arose from some misunderstanding or misfire. After all, she has received feedback in the past that she is prone to reacting dramatically and “making something out of nothing”.

Further suppose that, upon walking away from this experience, Wood finds herself believing of her boss “this person has wronged me”. She does not feel especially confident, and might not choose to assert her belief to others, since she does not understand what it is based on, and is certainly not prepared to cite reasons. Nonetheless, she cannot help forming the belief.

4.1 The Justification Question

Intuitively, Wood's belief is justified. But as with the Riley case, it does not meet the modified awareness requirement. As applied to Wood's case, the requirement would say that her belief is justified only if (i) There is something, X, that contributes to the justification of her belief B – e.g., evidence for B or a truth-indicator of B or the satisfaction of a necessary condition of B's justification; and, (ii) Wood is aware of X is such a way that she justifiedly applies to X the concepts of being something that contributes to B's justification and being something that could not easily attach to a stray hunch or an arbitrary conviction.

As in the Riley case, there are not many candidates for X, because there are not many things that Wood is aware of that might count as justification-contributors. She is aware of her feeling of intense discomfort, of the fact that she cannot help feeling that she has been wronged, and perhaps of the fact that she does not normally feel these ways after interacting with her colleagues. Does she justifiedly apply the concepts that condition (ii) requires to any of these? Take the first concept first. Wood may or may not apply the concept of being something that contributes to B's justification to any of these candidates for X. Although she is not prepared to cite these aspects of her experience as
reasons for believing, she may conceive of them as justification-contributors, and may be justified in doing so. For instance, it does not seem crazy to think that a persistent feeling that one has been wronged contributes some measure of justification to the belief that one has been wronged. But as in the Riley case, the reason that none of these candidates for X meets condition (ii) of the modified awareness requirement is that none of them are things to which she also justifiably applies the concept of being something that could not easily attach to a stray hunch or an arbitrary conviction. Indeed, the reason Wood is so unsure of herself, and unwilling to espouse her belief publicly, is precisely that she does not conceive of these candidates for X as things that could not easily attach to a stray hunch or an arbitrary conviction. After all, from her own perspective, she cannot rule out that these feelings that she does not usually experience in interacting with colleagues arose from some unrelated source and do not indicate anything at all about the content of her belief. So she does not see these as things that the belief has going for it, which thereby make it different from a stray hunch or an arbitrary conviction. In other words, her belief is vulnerable to the SPO. So as in the Riley case, the breed of accessibilism whose aim is to avoid the SPO does not have the resources to account for the idea that Wood's belief is justified.

Mentalism can account for the idea that Wood's belief is justified by appealing to the unconscious mental states that underlie the states she has access to. The thought is that subconsciously, Wood picked up on the badness in her interaction with her boss. She had some unconscious informational states that were responsive to that fact about the interaction, and those states were involved in causing the state of affairs in which she felt uncomfortable, and could not help feeling that she had been wronged. For the mentalist, those states can be sufficient to justify Wood's belief, even

69 There is a further question of whether, if Wood were to actually apply the required concept to X, she would do so justifiedly. It seems plausible that she would not be justified in conceiving of these as things that could not easily attach to a stray hunch or an arbitrary conviction. Her understanding that she cannot rule out that her feelings have some unrelated source, together with her knowledge of the feedback she has received about other reactions in the past, seem plausibly to act as a defeater of such justification. But again, spelling out what exactly it takes to justifiedly apply a concept to X is beyond the scope of my discussion.
though they are not accessible, and Wood is not aware of how they distinguish the status of her belief from that of a stray hunch or an arbitrary conviction.

As in the Riley case, the important thing is that mentalism is able to account for and accessibilism is not is the way in which Wood's belief is different from a stray hunch or an arbitrary conviction. And this has to do with the presence of her unconscious mental states. Like Riley's, Wood's unconscious informational states are the result of a type of expert perception, in the sense that adult humans are experts at picking up on when they are being treated badly or wronged by another person (at least in certain ways, or in certain contexts), even if they are not always aware of what it is they are picking up on. Indeed, we might think of the basis for Wood's belief as a type of social thin-slicing. 70

5. Implications for Our Social Epistemic Practices

We have seen that there is reason to think our justifiers need not be accessible in the sense that is required to avoid the SPO. Nonetheless, a rather demanding access requirement is built into some of our ordinary social epistemic practices. To see this, consider the concept of justification that is implicit in our practice of giving and asking one another for reasons for our beliefs. It is part of our ordinary social epistemic practice that we sometimes ask one another to cite reasons for our beliefs. If in response a person is unable to articulate reasons that seem to us sufficient to justify a belief, then in the context of this practice, we conclude that the belief is unjustified. 71

70 Providing thorough empirical evidence for the idea that adult humans are experts at picking up on when they are being wronged is beyond my scope here. What I suspect is roughly that evolution has gifted us with some mechanism or mechanisms that are pretty good at picking up on when another person is being threatening (which would be adaptive because it helps ensure physical safety) or asserting their power over us (which would be adaptive because it is reproductively advantageous to be attentive to facts and dynamics relating to social hierarchy when living in social groups). Wood's mechanisms are in good working order, but she's not interpreting her feeling of discomfort or danger as an indicator of a person being threatening or wrongly asserting power because she's been socialized to think this feeling does not indicate anything problematic, at least in the relevant context.

71 Indeed, Adam Leite seems to use the facts about this social epistemic practice as motivators for his view of justification. On the first page of his paper, he says, "...many people find it natural to think that if you can't justify a belief, then you aren't justified in holding it. Imagine, for instance, a case in which your interlocutor insists upon a surprising claim. He offers some reasons which manifestly fail to support the truth of his belief, and he has no satisfactory reply when this shortcoming is pointed out to him. His belief isn't justified, or so it would seem; he really shouldn't be so certain about
In a recent paper, Ian Werkheiser points out that this can be problematic, because the concept of justification embedded in this practice privileges those who are good at providing articulate, convincing-sounding reasons over those who in fact have good reasons for their beliefs. Fricker's sexual harassment case provides a nice example of a place where this distinction crops up. As I described her, the subject in this case has good reasons for her belief. Wood's belief that her boss has wronged her is based on her having picked up on the fact that her boss wronged her. However, she is not in a position to articulate the reasons for her belief, because articulability requires access, and she does not have access to the mental states that underlie and justify her belief. So, if an interlocutor asks her why she believes she has been wronged, she will not be able to offer a convincing response. As a result, her belief will not count as justified in the context of this ordinary epistemic practice.

What is even more worrisome about this example is that people who are in situations like Carmita Wood's are, in those contexts, some of the most likely to be pressed to articulate reasons in order to convince others to count their beliefs as justified. First, because Wood is in a situation of hermeneutic injustice, there is not a readily-available concept that would convince an interlocutor of the truth of her belief by subsuming her boss' behavior under a concept that makes obvious that it constitutes a moral wrong. Because of this, if she expresses her belief in the available terms, it is more likely to invite skepticism, particularly if her interlocutor has never had a relevantly-similar experience. And second, people in situations of hermeneutic injustice are already socially marginalized. As such, they will have diminished credibility in the eyes of many of their interlocutors right off the bat; so, the bar for what it takes to provide convincing-sounding reasons for their beliefs will be even higher.

This is a place where epistemic theorizing can help us to be critical of some of our ordinary

---

72 Werkheiser (2014).
social epistemic practices. Given that we have reason to think that having access to our justifiers is not necessary for justification, we should, at the very least, avoid engaging in practices that automatically treat a subject's belief as unjustified on the basis of her being unable to immediately provide convincing-sounding reasons when asked for them. This is not to say that there is an easy answer to the question of what our practices should be like. Our practice of giving and asking for reasons for our beliefs, of course, has an important function. Because we live in an epistemic community, and have an epistemic division of labor, we need a practice that helps us sort out when other peoples' beliefs are worth taking seriously and giving credence to. But one thing we do not want this practice to do is to end up sorting beliefs in a way that decides which beliefs are worth taking seriously as a function of how much social power the believer has. That would be to both compound social oppression, and, given what I have said about access and justification, to get the epistemology wrong. The worry is that, particularly when it operates in a context of hermeneutic injustice, this is what our current practice sometimes ends up doing.

6. Conclusion and Future Directions

The cases that I have discussed give us reason to re-think the received view that mentalism as a view about justification is not worth defending. I said that philosophers have offered a couple of different reasons for the received view. Conee and Feldman suggested that philosophers tacitly assume the mentalism and accessibilism are extensionally equivalent, since the kind of access on which internalist theories rely can reach all mental items that play a role in justification. We have now seen that mentalism can account for our intuitive judgments about these cases, while the breed of accessibilism we considered cannot. This is because access does not reach all mental items that play a role in justification. The other motivation for the received view was Bergmann's suggestion that mentalism gives up on the clearest and most compelling motivation for endorsing internalism in the
first place, which was that it is able to avoid the SPO. But given our discussion, we no longer have reason to view the SPO as an objection to a view of justification. If what I have said is right, then it does not follow from the subject's lack of awareness of how the epistemic status of her belief is different from that of a stray hunch or an arbitrary conviction that it is not a justified belief. More positively, the fact that mentalism can account for our intuitive judgments about these cases and accessibilism cannot is reason to think that mentalism has merit as a view of internalist justification. To be fair, it could in principle turn out that further epistemic theorizing convinces us that our intuitive judgments about the justification of Riley's and Wood's beliefs were wrong. But at the very least, my discussion puts the burden of reasonable argumentation back on the accessibilist, to provide a principled account of why a belief cannot count as justified unless the subject is aware of how its epistemic status is different from that of a stray hunch or an arbitrary conviction.

Another positive upshot of our discussion is that it suggests a direction for future inquiry. Although having access to the mental states on which one's belief is based does not seem to be necessary for justification, our discussion gives us reason to think that a belief's being based on appropriate mental states might be an important difference-maker vis-à-vis justification. The lesson of the Clairvoyant Riley case was that because his belief was not based on appropriate mental states, Clairvoyant Riley's belief was not different in the right way from a stray hunch or an arbitrary conviction. Thus, even though his belief was true, and formed by a reliable process, it was not justified.

At this stage, the suggestion that a belief's being based on appropriate mental states might be an important difference-maker for justification is a very modest proposal. First, it does not take a stand on what the appropriateness of appropriate mental states consists in. And it might turn out that this differs between cases like Riley's, and cases of non-perceptual beliefs, for instance. Second, the proposal does not go as far as to suggest that having a belief based on appropriate mental states is necessary for

73 To be clear, I am talking here about doxastic justification.
justification. Our discussion lends some support to the idea that it is necessary in Riley's case, but less
to the idea that it is necessary across the board. Perhaps, for instance, it would turn out to be necessary
only in cases of perceptual belief, or only in cases where the subject lacks access to her evidence. The
positive suggestion for future inquiry is that we test our judgments about justification against other
cases that vary along these lines in order to see how we might fill in this modest proposal.
Epistemic Oppression and Ameliorative Epistemology

1. Introduction

This paper has two goals. The first is to identify a particular type of wrong that is distinctly epistemic in kind. I call it *epistemic oppression*. The term it refers to a class of ways in which one can be oppressed in a manner that is distinctly epistemic.

Theorizing about phenomena like oppression, as well as about wrongs more generally, is typically the province of moral or political philosophy rather than epistemology. But there are several recent projects in philosophy that have aimed to understand how certain moral or political wrongs can also be distinctly epistemic in kind. These include Miranda Fricker's work on a concept that she calls *epistemic injustice*, and Kristie Dotson's work on *epistemic oppression*.\(^74\) They are both interested in injustice that occurs specifically within the realm of epistemic activity. She points out that our epistemic practices are played out by subjects who are socially situated, and this puts questions about power and social identity at center stage. Her aim is to take account of this, in order to shed light on the ethical aspects of our epistemic practices.

My project is a part of this same effort. For my purposes, Fricker's book on epistemic injustice provides a useful foil. Whereas Dotson is writing in a tradition that foregrounds historically specific, dangerously oppressive situations and develops epistemic theories designed to help women and people of color navigate them, Fricker is writing in a tradition that foregrounds abstractly described epistemic situations, and develops epistemic theories designed to highlight the politically relevant aspects of their structure. My starting point is more similar to Fricker's than to Dotson's, but I argue that Fricker does not go far enough. Her concept of epistemic injustice is too narrow, in that it overlooks or obscures the

\(^74\) Dotson uses the same term as I do, although what she means by epistemic oppression is different from what I mean by it.
ethical dimensions of some of our most common social epistemic practices. My first goal, then, is to carve out a concept of epistemic oppression that encompasses phenomena that Fricker’s notion leaves out, and that are central to our epistemic lives. In addition to bringing these important phenomena to light, I argue that the concept of epistemic oppression helps to explain what is normatively bad about instances of both epistemic oppression and epistemic injustice, by identifying certain badmaking features that the concept of epistemic injustice overlooks. In effect, the concept of epistemic injustice that Fricker identifies in her book is a species of the genus of epistemic oppression.

My second goal is to help explain why concepts like epistemic oppression and epistemic injustice have not gotten more air time in analytic, Anglo American epistemology. I think it has to do in part with our choice of philosophical methodology. Some of the dominant methodologies in contemporary Anglo-American epistemology build in assumptions that make it seem less natural to count phenomena like epistemic oppression and epistemic injustice as primary matters of epistemic concern. In particular, I argue that a certain type of conceptual analysis restricts the range of our concern in these unnoticed ways. Nonetheless, these phenomena ought to have a central place in our discussion. I suggest a methodology that I think can help rectify this, which I call ameliorative epistemology. 75

2. Epistemic Injustice and Epistemic Oppression

In Epistemic Injustice: Power and the Ethics of Knowing, Fricker characterizes epistemic injustice as a phenomenon that comes in two forms, which she calls testimonial injustice and hermeneutic injustice. So, a phenomenon counts as epistemic injustice if it is an instance of either testimonial or hermeneutic injustice.

75 The term ‘ameliorative’ comes from Sally Haslanger’s work on what see calls ‘ameliorative inquiry. See, for instance, Haslanger (2012).
For a different take on epistemic oppression, see Dotson (2014).
For a different take on ameliorative epistemology, see Bishop and Trout (2005).
Fricker also makes a proposal about the underlying principle that unites these two forms of epistemic injustice. Namely, epistemic injustice is a kind of wrong that is distinctly epistemic, because it consists most fundamentally in a wrong done to a person specifically in her capacity as a knower. By this, she means that it is a wrong done to a person in her capacity as either (1) a testifier; that is, a person who can share knowledge with others via testimony; or, (2) a possessor of knowledge; that is, a subject who has knowledge, or can come to know things about the world.

Given all of this, we can extract two definitions of epistemic injustice: a disjunctive definition, and a liberal definition. On the disjunctive definition, a phenomenon counts as an instance of epistemic injustice iff it is a case of either testimonial or hermeneutic injustice. On the liberal definition, something is an epistemic injustice iff it consists most fundamentally in a wrong done to a person in her capacity as a knower. In principle, then, the liberal definition leaves space for forms of epistemic injustice that are neither testimonial nor hermeneutic injustice, since there might be other things that consist most fundamentally in a wrong done to a person in her capacity as a knower.

Numerous responses to Fricker’s work have pointed out phenomena that, while not instances of testimonial or hermeneutic injustice, seem intuitively like things that we might want to include under the label of epistemic injustice. In some cases, the liberal definition helps vindicate this intuition, since the phenomena are naturally understood as additional ways of wronging a person in her capacity as a knower. But others are not as easily understood this way. This raises a question: Are there other distinctively epistemic ways of wronging a person, besides the ones that are captured by Fricker’s disjunctive and liberal definitions of epistemic injustice?

I will argue that there are. I think there is distinctly epistemic kind of wrong that is not a wrong to a person specifically in her capacity as a knower. I will argue that epistemic oppression is that kind of wrong. Fricker's concept of epistemic injustice is simply the knowledge-centric species of the genus
of epistemic oppression.76

In the next section, I will offer a definition of epistemic oppression. The underlying idea – akin to Fricker's principle that epistemic injustice consists most fundamentally in a wrong to a person in her capacity as a knower – is that epistemic oppression consists fundamentally in a wrong to a person in her capacity as an epistemic agent more broadly. So, on the way that I am using these terms, *epistemic agent* is not synonymous with *knower*.

This might seem wrongheaded, insofar as we define epistemology strictly, or narrowly, as the study of knowledge. But part of the motivation for articulating a broader notion of epistemic wrong is the more general thought that epistemology as a field has focused too narrowly on the concept of knowledge, and on belief as a component of knowledge, to the exclusion of other interesting epistemic phenomena. Part of what I aim to show in carving out a concept of epistemic oppression is that focusing on the epistemic agent as a knower deflects attention from other factors that guide deliberation and inquiry. And as Fricker herself points out, because these processes are often played out interpersonally, they are where the epistemic wrongs take place. This is one way in which epistemology's intense focus on the concept of knowledge fails to give us a complete picture of the normative significance of our epistemic lives.

Since we are focusing on the ethical dimensions of our epistemic lives, it is not just dimensions of epistemic normativity that we risk overlooking here, but ethical ones as well. In this context, then, epistemology's intense focus on knowledge - and thus on the epistemic agent as a knower - functions as an enabling condition for a certain kind of oppression. This is because it obscures the ways in which a person can be oppressed in her capacity as an epistemic agent more broadly.

Divorcing the notion of an epistemic agent from that of a knower might leave some feeling untethered. While there is not space for a complete articulation and clarification of an alternative

76 Thanks to Mikkel Gerken for suggesting this language.
principle for circumscribing the realm of the epistemic, or of epistemic agency, the basic picture is one on which the value of being an epistemic agent is not limited to or derivative of the value of being a knower. Indeed, there are other things that we do, qua epistemic agents, the value of which cannot be understood wholly in terms of their role in the pursuit of knowledge. Other valuable exercises of epistemic agency might include things like: gaining understanding; or, relatedly, coming into 'cognitive contact with reality'; determining which questions to open and when to close them; gaining true beliefs that are valuable independently of whether they count as knowledge; and, developing one's epistemic capacities and virtues. I will elaborate on some of these in §3. Although many of these exercises of epistemic agency are tied to particular theories in epistemology, the underlying picture that they collectively illustrate has considerable intuitive pull: the idea is that the value of our distinctively-human breed of rationality cannot be cashed out solely in terms of our capacity to bear the kind of relation to a proposition that is described by epistemology's best theory of what it means to know.

In what follows, I offer a definition of epistemic oppression. Since it draws on a broader notion of epistemic agency, epistemic oppression is broader than epistemic injustice in the sense that all instances of epistemic injustice will also count as epistemic oppressions, but not vice versa. In marking out the sort of phenomena my definition encompasses, I will identify some things that count as epistemic oppression but not epistemic injustice, as well as some phenomena that count as both, and some that have an ambiguous status. Ultimately, whether or not a particular phenomenon counts as an

---

77 For examples of pictures on which understanding is distinctively valuable, and not a species of knowledge, see for instance Kvanvig (2003), and Elgin (1996).

78 For Zagzebski, the goal of the intellectual virtues is “cognitive contact with reality”, which includes but is not reducible to knowledge. For instance, it might also include understanding, which, unlike knowledge, includes the comprehension of abstract structures of reality apart from the propositional. See Zagzebski (1996).

79 For more on this, see Friedman (2013).

80 See Baehr (2011) for a picture on which intellectual virtues are to be thought of as traits, or excellences, that contribute to their possessor's “personal intellectual worth”. My thought is that the value of developing these excellences need not be understood as derivative of the value of knowledge.
instance of epistemic injustice is a verbal issue. What is important is the fact that the concept of epistemic injustice that Fricker identifies is not the only kind of epistemic wrong that exists, since epistemic oppression is another such wrong. Adding the notion of epistemic oppression to our conceptual repertoire is important, first, because it helps shed light on instances of epistemic wrongdoing that we might otherwise have overlooked. If we focus solely on epistemic injustice, we will miss a wider class of phenomena that can have a profound effect on how our epistemic lives go, and in turn, on how our lives go more generally. Second, the concept of epistemic oppression is important because it helps us gain a fuller understanding of what is normatively bad about the phenomena that we are already aware of. The notion of epistemic oppression points us to the idea that even in clear cases of epistemic injustice, there are badmaking features that Fricker’s concept overlooks.

2.1 Defining Epistemic Oppression

What is epistemic oppression, and why is it a distinctly epistemic kind of wrong? Here is a preliminary definition:

One is *epistemically oppressed* when external conditions impinge on her flourishing *qua* epistemic agent in a way that results from an operation of social power, rather than being incidental to it.

There is a lot to spell out here. I will begin by saying what I mean by *social power*, and what I mean by *oppression*. My definition of *social power* follows Fricker's: *social power* is a socially situated capacity to control others’ actions. This capacity can be exercised by particular social agents, or it can operate purely structurally. In the case of purely structural operations of social power, the capacity to control others’ actions is disseminated throughout the social system.\(^8\)\(^1\) There is no particular agent doing the

---

\(^8\)\(^1\) Fricker gives the example of a social group that is disenfranchised in that they tend not to vote for reasons that have to do with a structural operation of social power. In that case, the capacity for control is not exercised by any particular agent, but nonetheless it is a social power, because it is dependent upon the context of a functioning social world with shared social institutions, social meanings, sets of expectations, \&c.
controlling; nonetheless, there is always a social group that is properly described as being controlled.

This definition of social power is tailor-made for talking about action rather than belief. But for our purposes, we can understand the relevant actions as exercises of epistemic agency. For instance, one might exercise social power by exerting control over an agent's belief-forming, or deliberative activities. Imagine a case of group deliberation in which one participant’s ideas are routinely ignored to such an extent that she herself comes to believe they are not worth taking seriously, and thus begins to discount them and rely on others’ ideas instead. What is operating in this case is a socially situated capacity to control exercises of the person’s epistemic agency; here, by controlling her deliberative activity.

Following Sally Haslanger, the concept of oppression that I am using points to the ways in which groups of individuals are systematically and unfairly disadvantaged within a particular social structure. I want to point to the ways in which people are thus disadvantaged within a particular social epistemic structure. The basic thought is that within a given social structure, or network of social relations, differences in social power and social authority routinely affect how we perform epistemic activities by systematically disadvantaging some at the expense of others, specifically in their capacity as epistemic agents.

Putting together these definitions of social power and oppression:

Epistemic oppression happens when and only when members of a group are subject to a systematic and unfair disadvantage within a social epistemic structure, resulting from the operation of a capacity to control exercises of their epistemic agency, thus systematically and

---

82 It is also important to Fricker that central cases of epistemic injustice are systematic, but she seems to set the bar for systematicity somewhat higher than I intend to for the purpose of talking about epistemic oppression. For her, systematic injustices are produced by prejudices that track the subject through different dimensions of social activity – economic, educational, professional, sexual, legal, political, religious, and so on. Importantly, she says, this gives us a sense of how epistemic injustice fits into the broader pattern of social injustice. While I agree that a phenomenon has greater structural social significance to the extent that it is systematic in Fricker’s sense, I will call something ‘systematic’ only to distinguish it from being merely incidental vis-à-vis social power and authority. And the effects of social power and authority can be domain specific in some cases, such that they do not track the subject through these other dimensions of activity.

unfairly limiting their capacity for epistemic flourishing.

There are a couple of things to emphasize about this definition. First, epistemic oppression is distinct from epistemic injustice. When someone epistemically oppressed, she is not necessarily wronged in her capacity as a knower, but she is necessarily wronged in her capacity as an epistemic agent more broadly, since her capacity for epistemic flourishing is unfairly limited. Since the value of epistemic agency is not limited to or derivative of the value of being a knower, a person can fail to flourish epistemically in myriad ways aside from being undermined in her capacity as a knower. A few of the examples that I will expand on in §3 are as follows: she might fail to form important true beliefs; she might be barred from participation in intellectual deliberation that aims at understanding; or, she might fail to develop epistemic virtues, to name a few.84

The second thing to point out about this definition is that where the label ‘epistemic oppression’ applies, it points to the fact that the phenomenon it picks out has both ethically and epistemically badmaking features. Epistemic oppressions are ethically bad by virtue of being oppressions. The label *oppression* here is not just a metaphor; rather, being oppressed in one’s capacity as an epistemic agent is a genuine kind of oppression, albeit one that targets epistemic agency in particular. And epistemic oppressions are epistemically bad because being epistemically oppressed limits a person’s capacity for epistemic flourishing.

To make more intuitive both what kind of wrong epistemic oppression is and how it is distinct from epistemic injustice, I will give some examples both of phenomena that count as epistemic oppression but not epistemic injustice, and of phenomena that might count as both. This will illustrate

84 It is worth addressing the question of why I chose to use the word *oppression* to define a notion separate from epistemic injustice, rather than simply arguing that Fricker’s notion of epistemic injustice ought to be broader. Terminologically, the word ‘oppression’ is a more natural choice than ‘injustice’, certainly for capturing the features that I am concerned about, and perhaps for Fricker as well. First, the word ‘injustice’, historically, is divided into two senses, of distributive and corrective injustice, and neither goes to the heart of what is meant by epistemic injustice or epistemic oppression. Second, the word ‘oppression’ is a more natural choice because it highlights the structural relations between social power and social identity, which are important both for Fricker and myself.
how the concept of epistemic oppression helps to both shed light on epistemic wrongs that we might otherwise overlook, and to explain what goes wrong even in cases where the relevant phenomena also count as epistemic injustice.

3. Illustrating the Wrong of Epistemic Oppression

3.1 Fricker’s Testimonial and Hermeneutic Injustice

Allow me to start with examples of Fricker’s concepts of testimonial and hermeneutic injustice, which make up the disjunctive definition of epistemic injustice. I will explain how these phenomena also count as instances of epistemic oppression. Later, I will give some examples that show that not all epistemic oppressions are epistemic injustices.

For Fricker, the primary type of epistemic injustice is testimonial injustice. It occurs when, in offering testimony, a speaker suffers a credibility deficit because of a prejudice on the part of the hearer, attaching to some aspect of the speaker’s social identity. 85 Karen Jones offers a striking case of this in a paper on the politics of credibility. 86 The example features a women named Fauziya Kassindja, who fled her native Togo in order to escape a forced polygamous marriage and female genital mutilation. Upon arrival in the United States, she gave herself up to the INS, claiming asylum as a refugee from FGM. The judge who heard the case denied Fauziya’s application for asylum, because he did not find her testimony to be credible: “I have taken into account the lack of rationality, the lack of consistency, and the lack of inherent persuasiveness in her testimony, and have determined that this alien is not credible.” 87 Jones hypothesizes that what went wrong here is that the judge had a prejudice attached to an aspect of Fauziya’s social identity (in this case, he was known not to be sympathetic to

85 Fricker (2009), p.28.
86 Jones (2002).
female asylum seekers), and so he assigned a lower plausibility rating to the content of her testimony than he would have absent this prejudice. Because of this, Fauziya’s testimony received less credibility than it deserved, given that it was not inherently irrational, inconsistent, or unpersuasive. The effect of this was that she was not able to communicate the relevant information via testimony in this instance. And as a result, she lost her case.

Testimonial injustice counts as a kind of epistemic oppression. In cases of testimonial injustice, members of a social group are subject to a systematic and unfair disadvantage within a social epistemic structure, resulting from the operation of a capacity to control exercises of their epistemic agency (namely, their offering of testimony) by not meeting it with uptake. And for reasons that Fricker does a nice job of articulating, this systematically and unfairly limits epistemic flourishing, since it undermines their ability to exercise the core epistemic capacity of communicating information via testimony.\(^{88}\) In Fauziya’s case, she is subject to a disadvantage, \(qua\) female asylum seeker, within a structure in which the judge is afforded a formal kind of epistemic authority that makes him the arbiter of what is reasonably believable in this context, such that he has the capacity to control these testimonial exercises of her epistemic agency.

Fricker’s second type of epistemic injustice is hermeneutic injustice, which is the injustice of having some significant area of one’s social experience obscured from collective understanding owing to a structural identity prejudice in the collective hermeneutical resource. More colloquially: a person is subject to hermeneutical injustice when, because she is in a position of social powerlessness, some experience of hers is not reflected in our collective vocabulary, and as a result, that experience is poorly understood by society at large and perhaps even by the subject herself.\(^{89}\) The classic example involves

\(^{88}\) It is worth noting here that if you think that agents can communicate information via testimony in cases where they only have justified beliefs (as opposed to knowledge), then Fricker needs a story about why instances of testimonial injustice in which the agent only has justified beliefs also consist most fundamentally in a wrong done to the agent in her capacity as a knower.

the experience of sexual harassment. Prior to the 1970's, the concept of sexual harassment did not exist in the United States.\textsuperscript{90} Thus, instances of sexual harassment were likely to be interpreted as something else, for which there was an available concept, like flirting or flattery. As a result, it was very difficult for women to understand their own experiences of being harassed, to describe those experiences to others, and to articulate the burdens and limitations that negotiating with these behaviors placed upon them in the workplace.

Again, this counts as epistemic oppression. These women were subject to a systematic and unfair disadvantage within a social epistemic structure, resulting from a capacity to control exercises of their epistemic agency. Here, the relevant exercises of their epistemic agency are not only testifying about their experiences of harassment, and being able to articulate the ways in which they imposed burdens and limits on their professional lives, but also fulfilling the conditions for knowing about the normative contours of those experiences in the first place, as well as having others know about and understand them. In this case, they lack the concept that would figure in the content of the knowledge that, for instance, “I am being sexually harassed”.

Fricker’s account of what is going wrong in these cases is accurate, so far as it goes, but there is more to say. Framing testimonial and hermeneutic injustice as epistemic oppressions adds to our understanding of what goes wrong, since it reveals that there is a distinctively epistemic kind of wrong taking place that does not simply consist in the subject’s being undermined in her capacity as a knower. The epistemic harm she is subject to is not exhausted by the fact that she is undermined in her capacity as a testifier and/or a possessor of knowledge. Rather, she is also undermined in her capacity as an epistemic agent more broadly. To name just a few further epistemic harms: The subjects in these examples might fail to form true beliefs about their experiences that are valuable independently of whether they count as knowledge; they might be conditioned by these experiences not to trust their

\textsuperscript{90} Fricker (2009).
instincts, and thus fail to develop certain epistemic virtues, or capacities for reliably forming true beliefs; they might even develop an implicit or explicit view that beliefs they cannot defend to the satisfaction of certain powerful others are objectively indefensible, and therefore unjustified. In §4, I will say more about why unfairly limiting a person’s capacity for epistemic flourishing in these ways in such a significant wrong, ethically as well as epistemically.

In what follows, I give four additional examples of epistemic oppression, and highlight the ways in which epistemic oppression is distinct from epistemic injustice. The first two examples do not count as epistemic injustices on either the disjunctive or the liberal definition. The others are more ambiguous in that they might be construed as epistemic injustices on the liberal definition. Nonetheless, I will illustrate how it is useful to understand them as epistemic oppressions, both to bring the phenomena to light in the first place, and to gain a better understanding of what is morally and epistemically bad about them.

3.2 Deliberative Injustice

Participating in the epistemic activity of inquiry often involves exchanging information via testimony. But it involves more than this as well. For instance, it can also involve floating ideas, jointly fleshing them out, asking questions that are intended to move a discussion forward, and considering various possibilities by evaluating their likelihoods.91

This happens not only when we are exploring ideas in the context of an intellectual conversation, but is a routine part of our daily lives. Two people who are jointly deliberating about something simple, such as how to get from Point A to Point B, will typically offer up suggestions, and jointly evaluate their costs and benefits based on factors such as what they judge to be the likelihood of

91 It is worth noting that Fricker’s concept could perhaps handle this if she were to give a much broader construal of what counts as testimony. But this is not the construal that is common in the literature on testimony, nor is it the one that Fricker works with in her book.
encountering various obstacles.

In a paper responding to Fricker’s work on epistemic injustice, Christopher Hookway describes a kind of epistemic wrong that befalls a person when she is unjustly treated as unable to participate in these kinds of deliberative activities with other people – activities of a sort whose content is intrinsically-epistemic.\(^92\) This wrong occurs when one person unjustly fails to regard another as competent or trustworthy in carrying out these deliberative activities. I will call this phenomenon *deliberative injustice*.

Deliberative injustice is a kind of epistemic oppression. The wrong here is different from epistemic injustice, because it undermines the person not in her capacity as a subject of knowledge, but as an epistemic agent more broadly, by undermining her very ability to participate in this kind of deliberation, which is an intrinsically epistemic activity. What’s more, this kind of epistemic activity need not even aim at knowledge. To illustrate this, consider the following example.

Imagine that students in a philosophy class in the northeastern part of the United States are asked to present in pairs on a reading they have been assigned. Each pair is tasked with presenting a summary of an article, raising some questions, and leading a discussion about whatever they decide are the muddiest points in the dialectic. Suppose one pair is composed of a student who has strong southern accent, and another who has an implicit or explicit view that people from the south are intellectually unsophisticated.\(^93\) Having each read the article, these two students meet to talk about how to understand and present the argument to their classmates, and what important issues it might raise about how to frame the relevant philosophical debate. During the course of their conversation, the southern student raises the sort of question that is intended not as a request for information, but as a contribution to continuing the discussion by suggesting a line of inquiry. Because of her bias, the other student either

\(^92\) Hookway (2012).

\(^93\) In the United States, a common bias associates accents from the southern part of the country with a lack of intellectual sophistication.
ignores the question, or interprets it as a request for information that is only loosely related to what the southern student asked.

In this exchange, the southern student is not being treated by her conversational partner as someone who can be a genuine participant in philosophical discussion. The other student’s response assumes that it is not worth trying make progress by carefully considering what the southern student has in mind, getting her idea into focus, and exploring it together. So, the southern student here finds herself unable to be recognized as a participant in the kind of intellectual deliberation that aims at an improved understanding of the philosophical debate under discussion.94

This is a case in which the southern student is epistemically oppressed. She is subject to a systematic and unfair disadvantage within a social epistemic structure, owing in this case to a prejudice associated with her accent. Her fellow student has the capacity to control exercises of her epistemic agency in this instance, since the southern student’s ability to engage in the epistemic activity of joint deliberation depends on her having a partner who takes her to be capable of doing so. This limits her capacity for epistemic flourishing, because it keeps her from participating in this kind of activity, whose content is intrinsically epistemic.95

Nonetheless the southern student is not subject to an epistemic injustice in Fricker’s sense, on either the disjunctive or the broad definition. First, this is not a case of hermeneutical or testimonial injustice. It is not hermeneutical injustice because the situation would not be corrected by adding a new concept to our collective hermeneutical resource. Nor is it testimonial injustice, even though it is structurally similar insofar as the student is excluded from discourse in a way that deprives her of

94 The structure of this example is inspired by an example from Hookway’s paper, though the content is different.
95 Another way in which this undermines the southern student’s capacity for epistemic flourishing is indirect – because she is unable to participate in this kind of activity, she lacks the opportunity to develop her epistemic, deliberative capacities in the ways that should would otherwise be able to. Even supposing that the student were not now competent at this kind of deliberative activity, she may have the capacity to become competent, and that capacity may go unrealized as a result of this kind of phenomenon.
epistemic flourishing. The difference is that the mechanism that effects her deprivation is not a credibility deficit attaching to her testimony in particular. Indeed, many of the southern student’s intended contributions to this type of conversation will not take the form of testimony at all, but might include things like suggesting a line of inquiry, leading the pursuit of a question, floating an idea in response to something her partner said, and the like. These aspects of the situation are not captured by saying that she is undermined in her capacity to share knowledge via testimony – indeed, she may never even intend to do so in this context. Hence, the case does not fit Fricker’s disjunctive definition of epistemic injustice.

It does not fit the liberal definition of epistemic injustice, either. On that definition, something is an epistemic injustice iff it consists most fundamentally in a wrong done to a person in her capacity as a knower, where that is either a testifier or a possessor of knowledge. We just saw that this is not a case of a wrong done to the student as a testifier. Nor is it a wrong done to her in her capacity as a possessor of knowledge. At the very least, we would be missing something important by trying to describe what goes wrong here solely in terms of the southern student’s being undermined in her capacity as someone who can come to know things about the world – the deliberation she takes herself to be involved in with her fellow student is not aimed at knowing some set of propositions, for instance, about what the most useful reconstruction of the argument is, how its pieces all fit together, and where work still remains to be done. Rather, the purpose of the exercise is for the students to jointly improve their understanding of the terms of a complex philosophical debate.

### 3.3 Distributive Epistemic Injustice

At the beginning of her book, Fricker points out that “Given how we normally think about justice in philosophy, the idea of epistemic injustice might first and foremost prompt thoughts about
distributive unfairness in respect of epistemic goods such as information or education."  

But she quickly sets this aside, since she is concerned with wrongs that are distinctly epistemic, and she believes that, "When epistemic injustice takes this form, there is nothing very distinctively epistemic about it, for it seems largely incidental that the good in question can be characterized as an epistemic good." Her thought here seems to be that although this kind of distributive injustice might be the sort of the thing that we intuitively or pre-theoretically associate with the term *epistemic injustice*, we do not stand to learn anything about a distinctively epistemic kind of wrong by looking at these examples. However, I will argue that we can learn something important by understanding these as examples of epistemic oppression, which is after all a distinctly epistemic kind of wrong.

Examples of distributive unfairness in epistemic goods like information and education abound. Some of the most striking ones are historical, such as when members of certain groups, like slaves, or women, were denied formal education. But this same kind of discrimination still happens less formally or explicitly, as when girls around the world miss school each month because of menstruation. Consequently, they fail to develop certain epistemic capacities and virtues to the extent that they would otherwise. These capacities include simple ones like the ability to read or write, as well as more complex ones, like critical thinking or scientific reasoning. As a result, they fail to develop epistemic virtues that require the actualization of such capacities. These might include things like fair-mindedness, carefulness and thoroughness in inquiry, interpretive sensitivity and discernment, intellectual courage, and rigor, to name a few.

This kind of distributive unfairness is an instance of epistemic oppression because in the relevant social context, girls are subject to a systematic and unfair disadvantage within the network of

---

97 Ibid.
98 In 1740, the first law again teaching slaves to write was enacted in South Carolina, as a result of fear that Black literacy would prove a threat to slaves’ dependence on masters. Cited in William Goodell (1853).
social relations in which they are prevented from going to school. The disadvantage results from the operation of a structural capacity to control exercises of their epistemic agency by preventing them from developing the capacities and virtues needed to exercise such agency in a way that is constitutive of their epistemic flourishing.

It is worth addressing the possibility that, although Fricker herself does not see it as such, this kind of distributive unfairness counts as epistemic injustice on the liberal definition, by consisting most fundamentally in a wrong done to these students in their capacity as knowers. After all, it may be that as a result of the fact that they lack the opportunity to develop epistemic capacities and virtues, the students are undermined in their capacity to come to possess certain items of knowledge, and even to share knowledge that they do have via testimony, since they are not seen by their interlocutors as well-educated. Nonetheless, to make the argument that the way in which these girls are disadvantaged consists most fundamentally in this particular aspect of their disadvantage would require something further. It would require a picture on which the value of developing one’s epistemic capacities and virtues is derivative of the value of actualizing one’s capacity as a knower.100

Next, I will illustrate two phenomena that we may be able to construe as epistemic injustices on the liberal definition. Nonetheless, it is useful to understand them as epistemic oppressions, in order to bring them to light in view of a more complete understanding of the normative contours of the situations, along both their ethical and epistemic dimensions. The first example is drawn from our ordinary epistemic practices. The second one, in contrast, is drawn from epistemological theorizing. It concerns the normative epistemic concept of justification.

3.4 Mis-calibrated Intellectual Self-Trust

---

100 This is not to say that those who argue in favor of knowledge first epistemology have not tried to develop such a picture. Ultimately, I do not think these attempts are successful, but there is not space to argue against them here.
In a paper responding to Fricker’s work, Karen Jones describes a phenomenon that she labels *mis-calibrated intellectual self-trust*. It occurs when one mistrusts her own cognitive capacities in a particular domain more than she ought to, given their level of reliability. This is something that often happens as a result of other kinds of epistemic injustice, such as testimonial, hermeneutic, and deliberative injustice.

Intellectual self-trust describes a stance that a person takes towards her own cognitive methods and mechanisms. It is socially created and sustained, such that it is porous to social power. Intellectual self-trust is created during our development as epistemic agents. When the results of our cognitive capacities are seconded by other people whom we trust, our self-trust is scaffolded and reinforced. But, given relations of social power, it is not always the case that this seconding reflects the *actual* reliability of our cognitive capacities; rather, it might simply reflect things like the kind of social identity prejudice that Fricker identifies. And this can give rise to mis-calibration of our intellectual self-trust, particularly in domains where members of a certain group are regarded as less credible than they in fact are. It might be that within a particular domain, a person’s interlocutors are frequently too critical or her beliefs and belief-forming processes, such that she herself comes to mistrust them. This likely happens, for instance with regard to beliefs that a person forms about her experience of microaggression. An important feature of microaggressions is that because they are in some sense minor (hence, ‘micro’), they are often carried off in a way that provides the offender some plausible deniability about whether any offense has taken place. Hence, when a person expresses a belief that she has experienced a microaggression, she may be told that she is simply being sensitive. If this happens repeatedly, she may come to mistrust herself, and to believe that she *is* merely oversensitive in this respect. The effect may be even stronger if she is in a situation of hermeneutic injustice, where she lacks the concept of a ‘microaggression’, which would help explain why her experience of feeling

---

101 Jones (2012).
degraded by the offender is legitimate.103

Just like there is a politics of credibility, there is a politics of self-trust. When a person’s intellectual self-trust gets mis-calibrated as a result, she is subject to an epistemic oppression. That is, members of the relevant group are systematically and unfairly disadvantaged qua epistemic agents by the fact that the mistrust themselves or their faculties more than they ought to given their actual reliability. And this results from the capacity of other people to exert a distorting influence on the development of the subject’s intellectual self-trust within a social epistemic structure in which their perspectives carry more weight.

This is a distinctly epistemic kind of wrong, which leads to distinctly epistemic kinds of harm. But these facts are not best or most fully explained by framing this as an instance of epistemic injustice, even on the liberal definition. This is consistent with the idea that mis-calibrated intellectual self-trust technically counts as an instance of epistemic injustice on the liberal definition. Indeed, it might be the case that as a result of mis-calibrated self-trust, a person is undermined as a testifier because she lacks confidence and so chooses not to testify. Or, it might be that she is undermined as a possessor of knowledge because her self-distrust keeps her from seeking knowledge in instances where she otherwise would have. Still, there is more to it than this. The normative contours of the situation are not fully captured by the idea that the harm consists most fundamentally in a wrong to the person in her capacity as a knower. She is also undermined in her capacity as an epistemic agent more broadly. Some examples of this are as follows. She might refrain from forming true beliefs that she otherwise would

102 Jones also points out that the pressure that epistemic injustice places on self-trust is not only causal, but is in fact a rational pressure. Because we count on one another for ongoing mutual epistemic calibration, there is a rational pressure for us to conclude that we have made a mistake when we cannot explain our divergent judgment.

103 In a separate paper, I argue for this idea that the mis-calibration of self-trust in response to social pressure is particularly likely to happen to people in situations of hermeneutic injustice. When these people form beliefs about the experiences that they cannot name, those beliefs are particularly unlikely to be vetted by other people, since we lack the conceptual resources to properly communicate about them. When a woman in the 1950's tries to communicate her belief that she has been wronged in an instance of sexual harassment, that belief is unlikely to be seconded by others. Her intellectual self-trust is likely to suffer as a result. At the result, this gives rises to what I call a kind of social gaslighting.
have. She might consider herself unable to participate – at least in certain domains – in deliberation that aims at understanding, of the sort described in §3.2. She might fail to develop epistemic virtues that she otherwise would have. She might fail to develop a worldview that leaves room for the idea that wrongs that are plausibly-deniable (such as the occurrence of microaggression) count as genuine wrongs.

Indeed, her very sense of what wrongs are plausibly deniable might also become worryingly liberal, so that ultimately she does not count an action as wrong unless she is able to immediately articulate a convincing explanation of why it was wrong. These aspects of the situation are not brought to light if we understand this as merely another case of epistemic injustice. The notion of epistemic oppression helps us understand how the person in this situation suffers a distinctly epistemic kind of harm that systematically and unfairly limits her capacity for epistemic flourishing in these ways.\footnote{Fricker does mention something like the possibility of mis-calibrated intellectual self-trust in her book. But she thinks of it as merely a secondary disadvantage that can attend repeated instances of testimonial injustice. Using the notion of epistemic oppression, we can understand it as a primary epistemic wrong.}

3.5 Oppressive Normative Epistemic Concepts

Our normative epistemic concepts themselves can be vehicles for epistemic oppression. Allow me to illustrate this with an example of an oppressive normative epistemic concept that is embedded in our everyday social epistemic practice.

In our current social epistemic practice, it is common to ask others about the reasons for their beliefs. If in response a person is unable to cite reasons that seem to us sufficient to justify a belief, then in the context of this practice, we conclude that the belief is unjustified. Ian Werkheiser points out that the concept of justification that is embedded in this practice privileges those who are good at providing articulate, convincing-sounding reasons over those who in fact have good reasons for their beliefs.\footnote{Werkheiser (2014).}

You might think that, far from being oppressive, this sounds like a perfectly good concept of
justification for our practice to rely on. But both intuitively and on other respected concepts of justification (reliabilism, for instance), this concept fails to count certain beliefs as justified, when in fact they ought to count.

Imagine a woman in a situation of hermeneutic injustice, who lacks the concept of sexual harassment, but forms the belief that she is being wronged by a colleague who is in fact sexually harassing her. Suppose that someone asks her why she believes that she has been wronged by this person. Since she lacks the concept of sexual harassment, which helps explain why the behavior is wrong, she will likely have difficulty articulating reasons that seem sufficient to justify her belief. Nonetheless, the belief may be formed by a reliable process: she may come to believe that she is being wronged because she is picking up on the fact of her being harassed, via a mechanism that reliably indicates when and only when this is the case. Given this assumption, reliabilism will count this woman’s belief that she has been wronged as justified. Even for those who are not reliabilists, I think there is an intuitive pressure to count her belief as justified. But it will not count in the context of the ordinary social epistemic practice I have described.

This constitutes a kind of epistemic oppression. The subject who is in this situation is subject to a systematic and unfair disadvantage within the context of a common practice inhering in our current social epistemic structure. In this context, her interlocutors exercise a capacity to control exercises of her epistemic agency in the sense that because her beliefs fail to count as justified (even though they should), she is barred from exercising epistemic agency in some of the ways that are constitutive of epistemic flourishing. Perhaps the most direct is that she is undermined in her capacity as a justified believer in a way that is structurally analogous to how one is undermined in her capacity as a knower

106 Indeed, Adam Leite’s view is that to be justified in believing some proposition, a person, must be able to develop and provide an adequate defense of her belief when asked to do so under appropriate conditions. See Leite (2004).
107 To make matters worse, this hermeneutic situation also makes it more likely that she’ll be asked the question in the first place, since it obscures the answer.
(i.e., a testifier) when she is subject to testimonial injustice, and unable to communicate knowledge via testimony. As a result of not being taken seriously as an epistemic agent who can form and communicate justified beliefs, the subject may also be barred from full participation in other interpersonal epistemic activities, such as joint deliberation. This kind of epistemic oppression may also cause miscalibrated intellectual self-trust, particularly if a person is subject to it repeatedly.

The best explanation of what has gone wrong here is not that the person is undermined in her capacity as a knower, but rather as a social epistemic agent more broadly. This is again consistent with the idea that this counts as an epistemic injustice on the liberal definition. Indeed, you might think that an important part of this phenomenon is that the person is undermined in her capacity as a possessor of knowledge, since justified belief is a component of knowledge. However, she is also undermined in that it limits her capacity to flourish as an epistemic agent in the additional ways I have described, and the concept of epistemic oppression brings these aspects of the situation to light.

Fricker’s notion of epistemic injustice fails to leave room for analyzing cases that do not consist most fundamentally in wrong to a person in her capacity as a knower. This is perhaps because her concept inherits epistemology’s intense focus on knowledge, and on articulating norms that, when followed, facilitate gaining knowledge, sometimes to the detriment of other interesting epistemic phenomena. Indeed, epistemology has concerned itself much more with knowledge, and with belief as a component of knowledge, than it has with other factors that guide inquiry. This fails to give a complete picture of the normative significance of our epistemic lives.

Broadening the focus would also serve Fricker’s stated goal of bringing more philosophical attention to the ethical dimensions of our epistemic lives. It would allow us to focus more on the epistemic process and less on the end result – or success state – be it knowledge or justified belief. As Fricker herself points out, epistemic processes are played out interpersonally, by subjects who are socially situated. Hence, this is where the ethical wrongs primarily take place. This gives us a partial
answer to question I will take up in the next section, of why concepts like epistemic oppression and epistemic injustice have not gotten more air time in epistemology. In what follows, I aim to go a bit deeper in exploring this question.

4. Epistemic Oppression and Ameliorative Inquiry

There has not been a tremendous amount of discussion within mainstream analytic epistemology of the ethical and political aspects of our epistemic practices. In this section, I will say something more about why this is, and will offer a suggestion as to how those of us who regard this as a failure might go about changing it.

Fricker offers an answer to the question of why these aspects of epistemology are under-discussed: “the theoretical framework of individualism and compulsory rational idealization that epistemology traditionally creates for itself makes it very hard to see how such questions might have a bearing on epistemology proper.” In what follows, I expand on this explanation by offering that in addition, the fact that this framework is so resilient has to do with the methodologies that we use in our epistemic theorizing. I will distinguish three different methodologies, and argue that epistemologists generally use the first two rather than the third. I will then argue that the third type of methodology – ameliorative inquiry – might help us solve the problem I have identified.

Sally Haslanger distinguishes three different things that philosophers might be doing when they engage in an investigation of some concept. The first is conceptual inquiry, which is used to seek an articulation or a clarification of a concept. In doing conceptual inquiry, we are interested in gaining a better understanding of the concept that we apply when we use a term in ordinary language. This proceeds by reflection on examples of cases where we use the term, with the goal of determining whether there are principles that unite them. This method has featured prominently in Anglo-American

epistemology, in helping us to articulate what we mean by normative epistemic concepts like justification and knowledge.

The second thing that philosophers might be doing when they investigate a concept is descriptive inquiry. This seeks to determine whether there is anything out in the world that our use of a concept tracks. It often proceeds by empirical investigation, looking out into the world to see whether there is something there that is being captured by our use of the term or concept. Such consideration of the phenomena can potentially help us to develop a concept that is more accurate than the one embedded in ordinary thought and talk. Within epistemology, this category is filled by projects in naturalized epistemology, which use this methodology to improve our understanding of normative epistemic concepts.

The third type of inquiry that Haslanger describes is ameliorative inquiry, which proceeds in two steps. In Step 1, it asks what legitimate purpose(s) we might have for having a particular concept at all. What work do we want the concept to do for us? In Step 2, it asks what concept we ought to have, given those legitimate purposes. What sort of concept can do that work?110

In our investigations of normative epistemic concepts, epistemologists have focused primarily on conceptual and descriptive inquiry. I will argue that this helps explain the fact that phenomena like epistemic injustice and epistemic oppression have been under-discussed. Conceptual and descriptive inquiry build in certain conceptions of what we should care about, and about what our goals ought to be in doing epistemology, and these assumptions preclude phenomena like epistemic oppression from counting as primary matters of epistemic concern.

I will argue for this claim by taking a closer look at conceptual inquiry, because descriptive inquiry (at least, the way that we usually practice it) also begins with conceptual inquiry. Let me

---

110 These different methodologies are discussed outside of political philosophy, too. For instance, Haybron (2010) makes similar methodological distinctions in a recent book, in order to consider different ways we might theorize about happiness.
I said that when we do conceptual inquiry, what we are looking for is an articulation or an analysis of our concept; that is, the concept that we take ourselves to be applying when we use the corresponding term. There are two ways that one could go about this – either by using reflective equilibrium, or not. When we are not using reflective equilibrium, we look for an analysis that both captures our ordinary notion exactly, in that it gives us necessary and sufficient conditions for something's being in the extension of the concept, and that is exhaustive, in that it does not admit of counterexamples. When we find an ordinary use of the concept that is not captured by our analysis, it counts as evidence that we have not yet gotten the analysis right.

If we are using reflective equilibrium, then the goal is to balance our intuitions about cases with our theoretical commitments. The aim is not necessarily to generate an exact and exhausting analysis of our ordinary concept. Still, we are doing conceptual inquiry. The difference is that we are looking to our use of the term or the concept not as a final court of appeals, but rather to fix the subject matter. We look to our ordinary concepts to figure out what it is that we value in epistemic matters, and we use this to fix ideas about what is in fact valuable epistemically.

For example, in examining our use of terms like justification and knowledge, epistemologists have observed that a core epistemic value that is embedded in these concepts has to do with having true beliefs and avoiding false ones. This has, to a large extent, fixed the subject matter in contemporary epistemology. Much of the discussion of these normative epistemic concepts revolves around how well they promote these epistemic values that are pulled from our ordinary concepts.

Thus far, I have been describing conceptual inquiry. But it is also the case that in doing descriptive inquiry, we use the values embedded in our ordinary concepts to fix the subject matter. Generally speaking, projects in naturalistic epistemology do not start by looking out into the world, and then using empirical information to construct normative epistemic concepts from the ground up.
Rather, they begin by using conceptual inquiry to fix the subject matter, and then they look to the world to revise our normative epistemic concepts in light of empirical information. The main way in which descriptive inquiry differs from conceptual inquiry is simply that the question of what concepts would best promote our epistemic values is answered in light of empirical information about what the world is like, and what our belief-forming mechanisms are like. So, in both conceptual and descriptive inquiry, the discussion is grounded in a consideration of how the analyses on offer promote the epistemic values embedded in our ordinary concepts.111

Given this, we are now in a position to issue the following challenge: Why should we assume that the epistemic values embedded in our ordinary notions represent all of the things and only the things that we should value in epistemology?112 Of course, this challenge becomes more interesting if we have positive reason to think that we should value something beyond the sort of epistemic goals embedded in our ordinary concepts. I will now argue that we do.

I see a couple of ways to make this argument. The first is to argue that because of the way our ordinary epistemic concepts were likely formed, for instance, in response to particular kinds of situational pressures, they may well be limited, or leave out some of the values that we should consider important. I suspect this argumentative strategy is onto something, but I will take a more direct route to the conclusion. The second way to show that our epistemic values should go beyond those embedded in

---

111 To some extent, of course, ameliorative inquiry must also use our pre-existing concepts to gain a foothold on what we are talking about. That is: our theorizing must in some way be answerable to pre-existing concepts or theories, lest we be rightly accused of changing the subject. Nonetheless, we need not proceed as if the only purchase we have on these notions comes from our antecedent commitments about them, and thus we need not use the degree to which the results of our theorizing conform to those commitments as a rubric for measuring our success. By starting with the normative question of what we want these concepts to do for us in the first place, the ameliorative project frames itself as one that has the potential to be more deeply revisionary.

112 In a paper on feminist values and normative epistemology, Haslanger issues an even more radical challenge. She asks, “Why should we place so much weight on our concept in the first place? If we take the primary task of epistemology to be a normative investigation of how we ought to reason, on what basis we ought to form beliefs, and more generally, about what is epistemically valuable, then there is something peculiar about a strategy that undertakes simply to describe our concept, or to discover the natural kind we ordinarily refer to. ...to suppose that what we value epistemically is what we ought to value epistemically is to leave the normative part of normative epistemology undone (Haslanger, 1999, p.466).”
our ordinary concepts is to provide a novel kind of answer to the question in the first step of an ameliorative inquiry. Doing so will illustrate how ameliorative inquiry can help broaden our conception of what is epistemically valuable, such that a discussion of epistemic oppression or epistemic injustice counts as doing epistemology proper.

For our purposes, we can formulate the question in Step 1 of ameliorative inquiry as follows: What legitimate purposes might we have for employing an epistemic framework? What kind(s) of work do we want our repertoire of normative epistemic concepts to do for us?

Certainly, one legitimate way of answering this question is by looking to the normative epistemic concepts we already have, and thinking about what purposes they serve for us, and the extent to which those purposes are legitimate. There is a lot to be gained from this. But we also have the option of giving a more revisionary answer. I want to suggest that the value of our epistemic practices is not limited to or exhausted by the value of the beliefs and knowledge that they produce. Rather, I think that there is an intrinsic value in our epistemic agency, owing to facts about the kinds of creatures we are.

Fricker alludes to this idea in her book, too, though she uses it in particular to defend the claim that there is an intrinsic injustice in being undermined in one’s capacity as a knower, because our capacity to be knowers is one side of the capacity that is essential to human value; namely, the capacity for reason. She points out: “We are long familiar with the idea, played out by the history of philosophy in many variations, that our rationality is what lends humanity its distinctive value.”113 For my purposes, the thought is that because the capacity for reason is essential to human value, and our epistemic agency is at least necessary to actualize that capacity (if not constitutive of it), our epistemic agency is intrinsically valuable. As a result, it is intrinsically valuable for us to be able to flourish *qua* epistemic agents. So, one legitimate purpose that we might have for employing a framework of

113 Fricker (2009), p.44.
epistemic concepts and norms is to promote our flourishing *qua* epistemic agents. And the concept of *epistemic oppression* has a natural place in such a framework, because in order to begin to think about how to address the kind of wrong that happens when someone is undermined in her capacity as an epistemic agent, we have to be able to name it, and to explain why it is a problem.

This also helps make clear why epistemic oppression is a distinctly-epistemic kind of wrong, and what is so bad about it. Epistemic oppression is a wrong done to a person in her capacity as an epistemic agent, where this is something of intrinsic value, and is the kind of thing capable of grounding normative epistemic concepts.
Bibliography


