Evidence and the Rationality of Belief

Anna Brinkerhoff, Brown University
This dissertation by Anna Brinkerhoff is accepted in its present form by the Department of Philosophy as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date______  

David Christensen, Advisor

Recommended to the Graduate Council

Date______  

Nomy Arpaly, Reader

Date______  

Joshua Schechter, Reader

Approved by the Graduate Council

Date______  

Andrew Campbell, Dean of the Graduate School
Curriculum Vitae

Anna Brinkerhoff’s philosophical interests lie primarily at the intersection of ethics and epistemology. She also has interests in feminist philosophy and the philosophy of religion. She received her M.A. in Philosophy from Brown University in 2017 and her B.A. in Philosophy from Pepperdine University in 2015.

Her published work includes:

The Promising Puzzle
Forthcoming – Philosophers’ Imprint

Death, Deprivation, and the Afterlife
Forthcoming – Philosopha

Why Epistemic Partiality is Overrated
2018 – in a special issue of Philosophical Topics (eds. Rima Basu and Mark Schroeder), co-authored with Nomy Arpaly

Review of Christensen and Lackey (eds.), The Epistemology of Disagreement
2018 – Analysis, co-authored with Tomás Bogardus
Acknowledgments

For helpful conversation, feedback, and support, I thank Zach Barnett, Emily Hodges, Chad Marxen, Arianna Falbo, Louis Gularte, Nina Emery, Tobias Fuchs, Nicolas Bommarito, Ginger Schultheis, Adam Pautz, and my partner Jason Bramburger.

Special thanks to David Christensen, Nomy Arpaly, and Joshua Schechter.

I am also especially grateful for my dad, who taught me to see gray where everyone else saw black and white, and for my mom, who showed me how to take on challenges (like writing a dissertation!) with grit, humor, and humility.
Table of Contents

Abstract ......................................................................................................................... p. 1
Dissertation Overview ................................................................................................. p. 2-4
Chapter 1: The Cognitive Demands of Friendship .................................................... p. 5-30
Chapter 2: The Promising Puzzle .............................................................................. pp. 30-59
Chapter 3: Prejudiced Beliefs Based on the Evidence ................................................. pp. 60-77
Chapter 4: Prejudiced Beliefs and the Moral Importance of Understanding.............. pp. 78-100
Beliefs bear a special relation to truth – beliefs, in some sense, seem to aim at truth – and so it is natural to think that there is an important link between truth and the rationality of belief (epistemic rationality). One view that captures this thought is evidentialism about epistemic rationality (evidentialism for short): it is epistemically rational for S to believe that p iff p is supported by S’s evidence. Although once widely accepted, evidentialism has recently come under attack along a variety of fronts. In particular, various philosophers have advanced accounts of friendship, promising, and prejudiced belief that challenge evidentialism. These challenges, although independent of one another, all rely on a common picture of the goodness of belief. Beliefs, the picture starts, have some important social, practical, or moral role: they enable us to be supportive friends, to make sincere promises, and to think without prejudice. But, the picture continues, there are cases when beliefs that are supported by the evidence cannot fulfill this other role: sometimes it is socially, practically, or morally good to not have a belief that’s supported by the evidence. So, the picture concludes, there is an important social or practical or moral dimension to the goodness of belief that is independent of, and sometimes in conflict with, belief being truth-aimed. Accepting this picture of the goodness of belief invites us to reject evidentialism. Everybody can agree that, if epistemic rationality is worth pursuing, it must promote good beliefs and discourage bad ones. But, if this picture is accurate, there are social, practical, and moral considerations that contribute to the goodness of beliefs. Epistemic rationality, we might think, must be sensitive to these other considerations.

The primary goal of this dissertation is to defend evidentialism. Towards this goal, I develop evidentialist-friendly accounts of friendship, promising, and prejudiced belief. Taken together, these accounts contribute to a systematic defense of evidentialism. They also indicate something important about the goodness of belief: the social, practical, and moral dimension of the goodness of belief is intimately connected with belief being truth-aimed.
Dissertation Overview

Beliefs bear a special relation to truth – beliefs, in some sense, seem to aim at truth – and so it is natural to think that there is an important link between truth and the rationality of belief (epistemic rationality). If belief aims at truth, the thought goes, then epistemic rationality is a guide to it: epistemic rationality is determined by truth-related considerations alone. One view that captures this thought is evidentialism about epistemic rationality (evidentialism for short): it is epistemically rational for S to believe that p iff p is supported by S’s evidence.

Although once widely accepted, evidentialism has recently come under question along a variety of fronts. In particular, various philosophers have advanced accounts of friendship, promising, and prejudiced belief that challenge evidentialism. These challenges, although independent of one another, all rely on a common picture of the goodness of belief, which looks like this:

Beliefs have some important social, practical, or moral role. Beliefs enable us, for example, to be supportive friends, to make sincere promises, and to think without prejudice. But there are cases when beliefs that are supported by the evidence cannot fulfill this other role. Sometimes it is socially, practically, or morally good to not have a belief that’s supported by the evidence. So, there is an important social or practical or moral dimension to the goodness of belief that is independent of, and sometimes in conflict with, belief being truth-aimed.

Accepting this picture of the goodness of belief invites us to reject evidentialism. Everybody can agree that, if epistemic rationality is worth pursuing, it must promote good beliefs and discourage bad ones. But, if this picture is accurate, there are considerations in addition to truth-related ones that contribute to the goodness of beliefs. Epistemic rationality, we might conclude, is sensitive to these other dimensions of the goodness of belief: social, practical, and moral considerations get a say in what is epistemically rationality to believe.

The primary goal of this dissertation is to defend evidentialism. Towards this goal, I develop an evidentialist-friendly account of friendship (chapter one), promising (chapter two), and prejudiced belief (chapters three and four). Each account is based on the rejection of one of the claims that make up the picture of the goodness of belief that underlies the relevant challenges to evidentialism. In the case of friendship, I
argue that belief does not have the social role that the picture says it has: that role belongs, instead, to attention. In the cases of promising and prejudice, belief does have the practical and moral roles the picture says it has but those roles are best fulfilled by beliefs supported by the evidence.

**The Cognitive Demands of Friendship**

In chapter one, I consider the view that friendship places demands on how we think about our friends. Simon Keller and Sarah Stroud both argue that being a good friend requires us to have positively-tilted beliefs about our friends even when they go beyond or against the evidence. Call this the doxastic account of the cognitive demands of friendship. Defenders of the doxastic account are committed to making a surprising claim about epistemic rationality: either that epistemic rationality is sensitive to considerations arising from friendship or that being a good friend sometimes requires one to be epistemically irrational. After considering both motivations of and worries for the doxastic account, I develop a new one: the attentional account. I argue that the attentional account can accommodate the various considerations that motivate the doxastic account while avoiding the problems that arise when our beliefs about our friends are peeled apart from our evidence about them.

**The Promising Puzzle**

In chapter two, I consider a puzzle about promising that it is tempting to solve by rejecting evidentialism. The puzzle arises from the thought that we should make a promise only if our belief that we will follow through is epistemically rational. But if that is right, and if it is epistemically rational to believe only what our evidence supports, then it seems that we should not make promises to do things our evidence suggests that there is a significant chance we will not do – things that many others, or we ourselves, have set out and failed to do. But surely that cannot be right! After all, these are some of our most important promises. One way to approach this puzzle is to reject evidentialism, and contend that it can be epistemically rational to believe that we will follow through even when this isn’t supported by the evidence. I argue against this approach, and any like it, that entails that it can be permissible to promise against the evidence. I go on to argue that, upon closer inspection, our evidence often does support the belief that we will do things that many others, or we ourselves, have set out
and failed to do. In these cases, promising is fine. But when our evidence really does not suggest that we will follow through, promising is not the right thing to do.

*Prejudiced Beliefs Based on Supporting Evidence*

In chapters three and four, I focus on a challenge to evidentialism that flows from considerations about prejudice. According to evidentialists, what is rational to believe is determined just by the evidence. So, assuming that prejudiced beliefs are irrational, evidentialists say that they must not respect the evidence. Recently, philosophers have been interested in cases of beliefs that seem to undermine evidentialism: these are beliefs that seem both prejudiced (and, thus, irrational) and based on of supporting evidence (and, thus, rational). For example, a server at a restaurant has statistical evidence that most Black diners tip less than average and then comes to believe that a particular Black diner will likely tip less than average. Several philosophers – call them revisionists – have appealed to these cases not only to reject evidentialism, but also to motivate moral encroachment. According the them, the server’s belief is not sensitive to the relevant moral considerations and so is irrational and prejudiced, even if it respects the evidence. In chapter three, I defend evidentialism from the challenge posed to it by beliefs like the server’s by arguing that the problematic intuitions can be explained away. In chapter four, I argue that the revisionist approach to cases like the server’s is too strong and that the evidentialist approach is, on balance, better. I do so by arguing that the evidentialist approach, but not the revisionist approach, can give a plausible evaluation of modified versions of the cases in question in which the beliefs are embedded in an understanding of relevant moral features of the believers’ socio-epistemic environment.

Taken together, these chapters contribute to a systematic defense of evidentialism. Perhaps more significantly, they tell us something important about the goodness of belief: the social, practical, and moral dimension of the goodness of belief is intimately connected with belief being truth-aimed.
Chapter 1
The Cognitive Demands of Friendship

Being a good friend is demanding. At the very least, friendship places demands on what we do. But perhaps being a good friend demands more from us than merely *doing* right by our friends. There is something natural to the thought that friendship places demands on how we think in addition to demands on what we do.

What, exactly, might friendship demand of us cognitively? The most prominent account of the cognitive demands of friendship is the doxastic account. Its defenders argue that friendship places demands on belief.¹ Specifically, they argue that friendship demands that we have positive beliefs about our friends. Sarah Stroud, for example, says that good friends should have “if not a blindspot, at least less than perfect vision where [their] friends’ sins and flaws are concerned [so that] the good friend’s set of beliefs is necessarily out of kilter.”² And Simon Keller says that good friends should have beliefs based on “considerations that have to do with the needs and interests of their friends, not with aiming at truth.”³

This entails a surprising conclusion about epistemology: sometimes, our beliefs should be sensitive things that do not bear on their truth. Friendship tells us, the thought goes, to readily believe the good and doubt the bad even when the evidence for each is equally strong; to believe our friend’s side of the story even when other sides are just as plausible; to believe the best of our friends even when the evidence supports something less flattering. For example, friendship might demand that I believe my friend is innocent even though the evidence tips in favor of guilt. Or it might demand that you believe that your friend will keep her promise to be on time even though you know she runs late more often than not.

The doxastic account is far from uncontroversial. Its rejecters raise a number of worries about it, and, of course, many of us will be hesitant to accept the claims about epistemology that it entails.⁴ That said, it does seem to get something important right: we should think in certain, perhaps positive, ways about our friends –

---

¹ I will focus on the doxastic account as articulated and defended by Keller (2004) and Stroud (2006), but Baker (1987), Hazlett (2013), Meiland (1980), and Keller (2018) advance similar views.
² Stroud (2006), p. 513
⁴ For discussions about worries for doxastic account, see Arpaly and Brinkerhoff (2018), Crawford (2019), and Kawall (2013). I will discuss several worries in section 3.
that is, friendship does issue cognitive demands even if they aren’t doxastic. My goal here is to characterize these demands in a way that avoids the epistemological conclusion and other worries that trouble the doxastic account while also accounting for the considerations that motivate it.

Towards that goal, I argue for the attentional account. On the attentional account, friendship places demands – not on belief – but on attention. The attentional account can accommodate the idea that friendship requires positive thinking by telling us to direct our attention towards the good when it comes to our friends, and away from the bad. It also renders intuitive verdicts in interesting cases – it enables us, for example, to praise those who dwell disproportionately on their friends’ virtues over their vices and criticize those who selfishly attend to their own needs when the needs of a friend are more pressing. But because friendship does not require beliefs on the attentional account, it is free of the surprising epistemological conclusion made by the doxastic account and avoids related worries based on considerations about belief and its norms. The attentional account has several additional advantages over the doxastic account because attention, unlike belief, is largely flexible: we can easily shift our attention away from the positive towards the negative and neutral and then back again. This feature of attention allows the attentional account to accommodate another important idea, one that does not fit as comfortably with the doxastic account: friendship sometimes requires impartial thinking in addition to positive thinking.

1 Friendship and Its Demands

Let’s begin with a general picture of friendship and its demands. The intended subject here is good friendship – the type of friendship shared, for example, by Elizabeth Bennet and Charlotte Lucas, Frodo Baggins and Sam Gamgee, Leslie Knope and Ann Perkins, and Harry Potter and Hermione Granger. A person with whom we share a good friendship is someone we might categorize as a best friend.⁵

Almost all accounts of friendship involve some sort of mutual care component. We care for our friends and for our friends’ own sake.⁶ Caring for our friends involves having certain desires and values related to them:

⁵ There may be many types of friends – work friends, hobby friends, friends who are children – to which the following considerations don’t apply.
a desire for them to flourish and achieve their ends, for them to be healthy and happy, to engage in meaningful relationships and projects. But mutual caring – though necessary for friendship – isn’t enough. As Aristotle noted, just because two people have good will for each other doesn’t mean that they are friends. Friendship requires something extra: satisfaction of the demands of friendship.

The list of demands might be thought to include a few fundamental demands and a lot of derivative ones. Fundamental demands are universal, applying to friendships across the board: help your friends achieve their important ends, for example, and promote your friends’ welfare. Derivative demands, on the other hand, generally apply but may vary across cultures and individual friendships. They tell us either how to go about meeting the fundamental demands, or how to go about fostering in us the desires and values involved in the mutual caring at the heart of friendship.

Some derivative demands are demands of action. They tell us what to do: lend a sympathetic ear to your friends when they need to vent, check on your friends when they are sick, acknowledge your friends’ birthdays, and celebrate your friends’ accomplishments. Other derivative demands – I take it – are demands of cognition. They tell us how to think. It’s plausible that friends ought to think positively about each another. So, when it comes to thinking about our friends, friendship might require positivity.

It is worth pausing here to note that “positive thinking” is meant to be neutral between the doxastic and attentional account – it does not implicate belief, and may be cashed out in terms of belief or attention. On the doxastic account, positive thinking about our friends involves positive beliefs about our friends, whether or not such beliefs respect the evidence; on the attentional account, positive thinking about our friends involves positive attention patterns – attending disproportionately to the positive features of friends, whether or not negative and neutral features are just as striking or salient or simply there to be noticed. Positive thinking is partial, but not because its constitutive thoughts are, on the doxastic account, necessarily counterevidential or,

---

7 *Nicomachean Ethics*, 1158a

8 Reasons for thinking that friendship requires positive thinking are explored in section 6. That said, there being cognitive demands of friendship to think positively is consistent with the idea, noted by Keller, that friendship does not demand positive thinking from all friends – for example, from perfectly virtuous friends. See Keller (2004) p. 334 and Keller (2018), pp. 27-30. Since cognitive demands of friendship are derivative, they can vary across individuals and cultures. What are taken to be common derivative demands of friendship apply in the majority of good friendships, but there may be some unique friendships in which they don’t apply.
on the attentional account, necessarily different from how we would otherwise direct our attention; rather, positive thinking is partial because it favors its objects without much regard for these other considerations.9

Interestingly, in many cases, cognitive demands seem to be parasitic on demands of action: in order to meet some demands of action, we must think in certain ways about our friends. Suppose that Jill has just landed her dream job. It’s plausible that friendship demands that her friend Jack act in certain ways in this situation – for example, by congratulating her and celebrating with her. But it seems problematic for Jack to publicly proclaim his delight while inwardly thinking that he wishes she hadn’t gotten it, jealous that she is more successful than him career-wise. This would, in Lindsay Crawford’s words, manifest “a bizarre split between a person’s outward and inward commitments to [his] friend.”10 Or as Keller puts it, “you want a friend who’s on your side, not one who’s good at faking it.”11 It seems, then, that when it comes to our friends, oftentimes partiality in action is to be matched up with partiality in thought, too.

Meeting the demands of friendship – both the demands of action and the demands of cognition – is partly constitutive of friendship. That being said, we need not meet all of the demands of friendship in order to count as a good friend. I take it that there is some minimal threshold of friendship demands that we must meet in order to count as a good friend: we are a good friend only if we generally meet these demands. But there is a normative element here, as well. We ought to meet the demands of friendship.

There are different ways the jump from the constitutive to normative claim can be made. It might be thought that having good friendships is a necessary component of a good life and we ought to do those things that are necessary for living a good life. Since meeting the demands of friendship is a necessary part of being a good friend and since being a good friend is, in turn, a necessary part of living a good life, we ought to meet the demands of friendship.12 Or, instead, it might be thought that, when we become friends with someone, we

---

9 When it comes to thinking about our friends, the extent to which positive thinking diverges from impartial thinking varies by friend. In the case of perfect friends, positive thinking about them might map squarely onto impartial thinking. It will be assumed here, though, that the vast majority of friends have plenty of imperfections so that, generally, positive thinking and impartial thinking about friends is significantly different.

10 Crawford (2019), p. 1579

11 Keller (2004), p. 335

12 Hints of this way of making the jump from the constitutive to the normative are found in Keller (2004) and especially Stroud (2006); Crawford (2019), p. 1579 explicitly lays it out.
effectively make a commitment to her to act and think in the ways constitutive of good friendship. Because we ought to follow through with our commitments, we ought to act and think in the ways constitutive of good friendship.

To sum up: friends care for each other, which involves them having certain desires and values regarding one another. Over and above mutual caring, friends must meet the demands of friendship. Some of them are fundamental demands, others derivative. Of the derivative demands, some are demands of action and some are demands of cognition. But what, exactly, does friendship demand of us cognitively?

2 The Doxastic Account

Let’s first consider the account of the cognitive demands of friendship that is already on offer. The doxastic account says that friendship gives rise to reasons for belief – specifically, reasons for us to have positive beliefs about our friends – and friendship demands that we take these reasons into account. This involves both partial doxastic practices and outcomes. To have partial doxastic practices when it comes to our friends is to act in ways that put us in a good position to form positive beliefs about them – for example, we might consult some sources of information and ignore others, seek out strengths of our friends’ work while overlooking weaknesses, and construct charitable interpretations of their behavior along with rosy portraits of their character. To have partial doxastic outcomes when it comes to our friends is just to have those positive beliefs: our beliefs about our friends should favor them and cast them in a good light, inclining us to admire them and like them and spend time with them. On the doxastic account, friendship requires both partial doxastic practices and outcomes, but the outcomes are more central – the practices are taken largely as means to the outcomes.

It’s crucial to note that doxastic outcomes demanded by friendship on the doxastic account will not always be supported by the evidence. Friendship does not tell us, on the doxastic account, to have beliefs only about the positive features of our friends. Rather, friendship tells us that the beliefs about our friends that we naturally

---

13 At its core, the doxastic account says that friendship gives rise to reasons for belief which friends ought to take into account; but, as it’s developed and discussed in the literature, the doxastic account is usually specified to say that these are reasons to have positive beliefs. See Stroud (2006), Kawall (2013), Crawford (2019), and Arpaly and Brinkerhoff (2018). Keller (2004) claims that friendship gives us reason to have beliefs that serve “the needs and interests of [our] friends” (Keller 2004, p. 330), but his primary focus is on positive beliefs that so serve.

form during the course of our friendships should, generally, tilt positive, whether or not such positivity to is supported by the evidence. Friendship, then, will sometimes require beliefs that go beyond, and even against, the evidence. This should be expected: after all, the reasons for belief flowing from friendship have to do with the needs and interests of our friends and the nature of friendship, while the reasons for belief flowing from evidence have to do with truth – it would be odd if they were always in sync.

So that’s the doxastic account of the cognitive demands of friendship. Why think it’s true? Perhaps for many, it just seems intuitive that friends ought to think positively about each other. But for those who lack this intuition, or for those who wonder why “think positively” should be cashed out in terms of belief, the defenders of the doxastic account offer a number of arguments in its favor. Keller, for example, suggests that truth of doxastic account is evidenced by familiar platitudes of friendship: “good friends believe in each other; they give each other the benefit of the doubt; they see each other in the best possible light.”

Another argument is that doxastic partiality is simply a natural extension of the partiality in action that friendship requires. Just as we ought to act in ways that favor our friends, so, too, ought we believe in ways that favor them. Of course, the demands of friendship are constrained by reality. Friendship does not demand the unrealistic or overly-taxing, in action or belief – as Stroud writes, “the good friend does not flatly believe the manifestly false, or refuse to believe the incontrovertible.” But friendship does demand some partiality in belief as well as action.

Yet another argument looks at how we adjudicate whether φing, for any φ, is a demand of friendship: we can do so by asking, from the first-person perspective, whether we would want our friends to φ and, from the third-person perspective, whether φing reflects well on someone as a friend. If we answer affirmatively to both, then that is a good indication that φing is a demand of friendship. It strikes some defenders of the doxastic

---

15 Perhaps the most compelling of these arguments are the ones for why friendship demands positivity – see especially pp. 338 – 344 of Keller (2004) and pp. 510-512 of Stroud (2006). I consider some of these arguments in section 7, although I consider them as arguments for why friendship demands positive thinking, as opposed to positive beliefs. For the sake of brevity, I will not also consider those arguments here.


17 Hazlett (2013), for example, writes on p. 93, “[W]hat makes the domain of belief special, that would insulate it from the kinds of partiality that seem required in the domains of action and motivation?... There is no reason to expect that. So partial love requires partial belief.” Although Hazlett’s point is about love, generally, it’s easy to see how it applies to the love of friendship, specifically.

18 Stroud (2006), p. 516
account that, upon reflecting on the nature of friendship along with particular cases of friendship, many of us will want our friends to generally have positive beliefs about us; many of us will also conclude that having positive beliefs reflects well on someone as friend.\footnote{This seems to be Stroud’s general strategy for defending the doxastic account. See pp. 501-502 of Stroud (2006). Arpaly and Brinkerhoff (2018) use this same strategy to argue against the doxastic account.}

In addition to these arguments, defenders of the doxastic account lean on particular cases of friendship and intuitions regarding what a good friend would believe in those cases. These cases are worth considering carefully since they feature prominently in defenses of the doxastic account. Let’s consider two of these cases, one from Stroud and one from Keller.

**Troubling Story** Ted goes to a bar after work with some of his coworkers and overhears a third party tell a troubling story about his friend, Emma. According to the third party – who is in the same Russian Literature class as Emma – Emma is constantly speaking over other students and interrupting the professor. Ted is hearing this story about Emma for the first time, and doesn’t know whether it’s true or false. The storyteller clearly disapproves of Emma. She has formed an unfavorable opinion about not only her troubling behavior (how rude!) but also her overall character (what a jerk!).\footnote{This case is adapted from p. 504 of Stroud (2006).}

**Poetry Reading** Rebecca is scheduled to read her poetry at a local coffee shop. Rebecca is nervous but determined to go through with the reading because she knows that a literary critic whom she wants to impress will be in attendance. She invites her friend Eric to the reading. Eric is surprised at the invite – he hadn’t known that Rebecca writes poetry, and isn’t familiar with her work – but happily agrees to come. As it happens, Eric has heard poetry read at this same coffee shop numerous times, and it’s almost always mediocre – almost everything that he has had heard read there isn’t of the caliber to impress a literary critic. He suspects curator of the readings simply has poor taste in poetry.\footnote{This case is from pp. 331 – 334 of Keller (2004).}

Defenders of the doxastic account can say something similar about both cases. In order to be a good friend, they might say, Ted and Eric would each meet certain demands of friendship. Ted would cast Emma in a good light and stand up for her in public. Eric would support and encourage Rebecca, and aim to see special value in her projects. In order to meet these demands – at least in a sincere way that doesn’t result in the ‘bizarre split’ discussed above – Ted and Eric must have certain beliefs about their friends. In fact, if they believe as friendship requires, their beliefs would be systemically different from those of someone who is listening to the same story or poem as a detached observer. They must assess the same data differently, drawing different inferences and conclusions.
For example, consider what Stroud says about Troubling Story.\(^2\) She suggests that Ted would be more confident in more charitable (and perhaps less obvious) interpretations of Emma’s reported behavior and less confident in less charitable interpretations. He might believe, say, that her behavior arises from passion about Russian literature as opposed to rudeness. If he can’t find a way to put a positive spin on her behavior, and is forced to admit that Emma acted rudely, Ted might instead “relegate [his] attribution of a character flaw…to an obscure corner of [his] portrait of her” – believing, perhaps, that rudeness isn’t an especially prominent part of Emma’s character, and is outweighed by her ardor and intellect.

Now consider what Keller says about Poetry Reading.\(^3\) Keller concedes that Eric would likely have accurate beliefs about the caliber of Rebecca’s poetry if it’s especially good or especially bad. But, he says, there are some cases – cases in which its quality falls somewhere between the two extremes – in which Eric would believe that Rebecca’s poetry made a good impression on the literary critic, even if he would have concluded differently had the same poem been read by a stranger. In these middle-of-the-road cases, in order to be a supportive friend, Eric would believe that Rebecca’s poetry is likely to have impressed the literary critic even though the evidence tips in favor of its being unimpressive.

Keller also argues that, as a good friend, Eric would have certain beliefs \textit{before} he hears the poetry. In order to try and see value in projects that are important to his friend, Eric must either believe that Rebecca’s poetry will likely be good, or, at worst, suspend judgment on the matter. Eric mustn’t believe, Keller says, that Rebecca’s poetry will be mediocre and unlikely to impress the literary agent before he even hears it – even though he has lots of evidence that any poetry read at this particular coffee shop is going to be mediocre and, we’re assuming, little evidence about Rebecca or her work to suggest that she is likely to be an exception.

Troubling Story and Poetry Reading seem to lend support to the doxastic account: it seems – at least, to those who are sympathetic to Stroud and Keller’s respective assessments – that friendship requires the friends in question to have positive beliefs and some of those beliefs (in at least some versions of the cases) are not supported by the evidence. Friendship requires Ted to have beliefs that enable him to cast Emma in a good

\(^{2}\) Stroud (2006), pp. 506 - 509
\(^{3}\) Keller (2004), p. 332 - 333
light and defend her publicly with sincerity, and friendship requires Eric to have beliefs that enable him to sincerely support Rebecca and see value in her projects – even if the details of the cases are filled out so that such beliefs are sunnier than Ted’s or Eric’s evidence permits.

3 Against the Doxastic Account

Now that we have considered some motivations for the doxastic account, let’s turn to some worries that trouble it. The first worry is that the doxastic account poses an uncomfortable tension between friendship and epistemic rationality: friendship tells us to base our beliefs on some things that don’t bear on their truth while epistemic rationality tells us the opposite. In the cases above, Ted and Eric may seem to be better friends by believing against the evidence. But drawing a general principle from these particular verdicts leaves the defender of the doxastic account with an uncomfortable claim: in at least some cases, and perhaps in general, the epistemically rational friend is a worse friend.

To bring this worry into focus, it’s helpful here to consider Arpaly and Schroeder’s case of a parent who worries excessively about her child. They point out that such excessive worry is often an indicator that the parent is caring enough about her child, not too much. Even so, it’s just an indicator: we shouldn’t criticize a parent who is a moderate worrier as long as she cares enough, and we shouldn’t urge her to worry more. It seems to me that something similar holds for friendship: irrationally positive beliefs often indicate that a friend is caring enough about his friend, not too much. Just as parental care can naturally give rise to excessive worry, so, too, the mutual caring between friends can naturally give rise to positively-tilted irrationality – perhaps, in part, because those we care deeply about tend to be our favorites, and as Arpaly and Brinkerhoff put it, “nothing is more human than… ‘not being objective’ about one’s favorite things.” Even so, irrationally positive beliefs

---

24 At least, this is what epistemic reasonably tells us according to standard theories of epistemic rationality which say that it’s determined only by evidential and other truth-related consideration. That said, it’s worth noting that, in her discussion of the doxastic account, Stroud (2006) considers non-standard theories of epistemic rationality that make room for non-evidential considerations like those arising from friendship. According to these theories, counterevidential beliefs required by friendship aren’t necessarily epistemically irrational. See Rinard (2017, 2019) for one such theory. Basu (2019), Basu and Schroeder (2019), Bolinger (2020), Moss (2018), and Schroeder (2018) all advance related theories according to which epistemic rationality is sensitive to moral considerations.


26 Arpaly and Brinkerhoff (2018), p. 49
about our friends are just an indicator of such caring – caring does not always yield irrationality – and what matters here is care, not irrationality. So we shouldn’t criticize the perfectly epistemically rational friend as a bad friend and we shouldn’t urge him toward epistemic irrationality, so long as he cares enough. But, on the doxastic account, we must.

Another worry is that the doxastic account implies that friendship produces obligations to have certain beliefs even though it’s plausible that we typically don’t have the type of control over belief that is necessary for such obligations. It’s long been widely accepted that having a moral obligation to φ requires that we have direct voluntary control over φing – at least when it comes to morality, ‘ought implies can.’ If obligations of friendship are relevantly similar to moral obligations, or if obligations of friendship are moral obligations themselves, then ‘ought implies can’ is in tension here with the fact that, typically, we do not have direct voluntary control over our beliefs about our friends. In other words, we cannot believe at will the things that friendship demands. So, on the doxastic account, the cognitive ‘oughts’ of friendship – namely, to have partial doxastic outcomes – are ‘oughts’ that we cannot typically satisfy. This leads us to worry that the doxastic account is incorrect or, at best, insignificant. If the doxastic account maintains that friends often have an obligation to believe in certain ways when, in fact, friends do not have such an obligation in cases where they lack direct voluntary control over their beliefs – which is, at least, a large majority of cases – the doxastic account is incorrect; if the doxastic account were revised to say that friends are obligated to have certain beliefs only in cases where they do have direct voluntary control over their beliefs – which is, at most, a small minority of cases – the doxastic account becomes largely insignificant.

The final worry is the most interesting – and also the most troubling for the doxastic account. It arises from an idea that even defenders of the doxastic account accept: sometimes, good friends should think

---

27 Doxastic involuntarists and voluntarists alike can accept this as fact. Doxastic involuntarists argue that there are no cases in which we have direct voluntary control over our beliefs; doxastic voluntarists argue that we have direct control over our beliefs in just some cases, and so may agree that, typically, we don’t

28 Importantly, the claim that, because of a lack doxastic control, there are no oughts of friendship on belief is consistent with the claim that, despite a lack of doxastic control, there are epistemic oughts on belief. Some philosophers argue that, given the nature of belief, epistemic oughts are special and do not imply can, even if other types of oughts do – see, for example, Feldman (2000, 2001). Whether or not there are epistemic oughts on belief does not bear on whether there are oughts of friendship on belief.

29 Keller (2004), p. 334
impartially about one another. Of course, one of the intuitions that the doxastic account is meant to capture is that we should think positively of our friends. In this vein, defenders of the doxastic account have said that friendship tells us to put a positive spin on our friends, to think well of them, to believe the good things and doubt the bad. But, there seem to be many reasons that flow from friendship to have a clear, untinted picture of our friends’ character, warts and all – at least some times and in some cases.

Very generally, there is something desirable about a friend who is aware that, say, her friend is pessimistic, or nit-picky, but loves him anyway, and still wants to be his friend. And there is something deeply undesirable about a friend who needs to be sheltered from the fact that her friend is a workaholic in order to find him loveable, or a friend who, upon realizing her friend is less intelligent than she once thought, is less inclined to be his friend. As Susan Wolf writes, “the best love…sees the beloved…clearly and fully, ‘as she really is,’ a love that sees the beloved’s faults and weaknesses as well as her virtues and strengths, and loves unreservedly nonetheless.”

In addition, there are plenty of specific reasons our friends might want us to think impartially about them, flaws and all. Perhaps friends should help each other become the best versions of themselves they can be – and we need to know of our friends’ flaws in order to help them overcome them. Perhaps, too, being a good friend involves giving good advice – and giving good advice often requires that we have a full picture of the person whom we’re advising. I might need to recognize that my friend is terribly disorganized, for example, in order to counsel against her opening a small business. And perhaps friendship demands that we prevent our friends from committing moral errors – something which would require us to know of their vices. You might need to know of your friend’s impatience, for example, in order to talk her down from lashing out at the overwhelmed server when the food is taking too long.

These considerations suggest that friendship calls for impartial thinking in addition to positive thinking, and this is where the doxastic account starts to falter. The doxastic account would work on the assumption that friendship issues a universal injunction to think positively – then it would simply say that friendship requires us

30 The quote is from p. 380 of Wolf (2014), who appeals to Iris Murdoch’s concept of “loving attention” to argue against the view that love casts its object in a positive light. Kawall (2013) makes a similar point on p. 357.
31 Arpaly and Brinkerhoff (2018) make a similar point about advice on p. 43.
to have a consistent set of positive beliefs about our friends: to believe that they are innocent and especially virtuous, that their vices aren’t so bad, that their work is superior, and that they will succeed, even if such beliefs don’t fit the evidence. The doxastic account would also work if it were assumed instead that friendship issues a universal injunction to think impartially – then it would simply say that friendship requires us to have a consistent set of rational beliefs about our friends: to believe in accordance with the evidence that they may be guilty, that they have some virtues as well as some vices, that their work is mediocre, and that their chances at success and failure are about equal.

But the doxastic account can’t easily handle a picture of friendship that requires some impartial thinking in addition to positive thinking. On the doxastic account, this would mean that friendship sometimes requires rational beliefs about the very things that, most of the time, friendship requires (sometimes irrationally) positive beliefs. So a good friend, on the doxastic account, must switch up her beliefs depending on the case – in one case believing that her friend is lazy and the next believing that he isn’t lazy but merely laid-back; in one case believing that her friend will get into her dream law school and the next believing that the odds are against her; in one case believing that her friend’s paintings are exceptional and the next believing that they are rather ordinary. Or else, a good friend, on the doxastic account, must hold grossly inconsistent beliefs, believing, for example, that her friend is lazy and while also believing that he’s not lazy. Either way, if friendship requires both positive thinking and impartial thinking, then, on the doxastic account, friendship in effect requires unsustainable doxastic gymnastics – doxastic switcharoos and gross inconsistencies are not only psychologically infeasible to sustain, but also, if sustained, would foster doxastic instability, thus preventing the good friend’s beliefs about her friends from effectively guiding her actions and grounding her grasp on reality.

4 The Attentional Account

Perhaps none of the three worries discussed is decisive against the doxastic account, and not everyone will find each worry troubling.³² But, especially when considered together, they mount a significant case against it and,

³² With regard to the first worry, some may simply be unbothered by a tension between friendship and epistemic rationality, or, even if bothered, argue that we should accept that there is a tension – just as some have argued that we should accept that there is a tension between friendship and morality. With regard to the second worry, some may argue that we do have the relevant sort of voluntary control over our beliefs – see, for example Shah (2002) and Steup (2016).
at least, give us reason to pause and look for alternative accounts of the cognitive demands of friendship. I want to suggest such an alternative account that is inspired by a move made in the literature on modesty.

Much of the contemporary discussion on modesty centers on the idea that modesty is a virtue of ignorance, a view advanced by Julia Driver. According to Driver, to be modest is to underestimate or be ignorant of one’s own good qualities. For example, imagine Jimi Hendrix saying, “Oh, I’m alright at playing the guitar.” Driver would say that Hendrix is modest only if he is ignorant of the extent of his musical talent. Some philosophers – uncomfortable with a view that roots a virtue in ignorance – offer alternative views on which modesty allows (and, some, necessitates) that the modest person have rational beliefs about his own good qualities. Nicolas Bommarito argues that the debate about what beliefs modesty forbids or permits is all misguided. Modesty isn’t about having or lacking any sort of belief, ignorant or rational, Bommarito argues: it’s about certain patterns of conscious attention. Specifically, he argues that modesty requires that we direct our attention away from our good qualities and their value, and toward external factors that aided in their development. A move similar to the one Bommarito makes in the literature on modesty can be made here, giving us the attentional account of the cognitive demands of friendship: friendship demands that we attend in certain ways to our friends – that we dwell on certain features, entertain certain possibilities, bring to mind certain memories, acknowledge certain qualities, and ignore others – not that we have certain beliefs about them.

The intuitions that motivate the doxastic account can be accommodated by the attentional account. Consider the general intuition that we ought to think positively about our friends. On the attentional account, we think positively our friends not by having positive beliefs about them but by, generally, directing attention towards the positive things – towards the good and away from the bad, towards the things about them that

---

33 Driver (1989)
34 For a sample of these alternatives to Driver’s view, see Statman (1992), Maes (2004), Raterman (2006) and Brennan (2007).
35 Bommarito (2013), p. 103
36 As indicated, attending is a multifaced cognitive activity; we may attend to something by dwelling or focusing on it, bringing it to mind, contemplating it, entertaining it, acknowledging it, and observing it, among other things.
make us smile, towards their successes and virtues, and away from their failures and vices. Holding fixed Jill’s beliefs about Jack, we might say that Jill thinks positively about Jack if she focuses on his compassion and how he regularly volunteers at the local soup kitchen more than his flakiness and how he regularly backs out of Friday night plans; we might say she thinks less-than-positively if she focuses on his flakiness and backing our more than his compassion and volunteering.37

In addition to accommodating this general intuition, the attentional account can also accommodate intuitions about the particular motivating cases. In Troubling Story, Ted might attend to interpretations of Emma’s alleged behavior that are different and more charitable than the one assumed by the storyteller, and he might attend to Emma’s positive traits that could explain or outweigh it. By speaking outwardly what is privately on his mind, Ted sticks up for Emma publicly without causing a ‘bizarre split’ between his inward and outward orientation towards her. In Poetry Reading, before the reading, and in order to make an effort to see value in Rebecca’s projects, Eric might direct his attention toward the power of poetry, towards Rebecca’s bravery in doing something nerve-wracking, and towards her characteristics – creativity and quirkiness, say – that incline him to believe that she may write good poetry, after all. As he approaches her afterwards with an encouraging word – “Good job, Beck!” – Eric might keep in mind the admirable parts of the poem and ignore the forced rhymes and the cheesy metaphor in stanza five. Like Ted, Eric supports Rebecca while avoiding the ‘bizarre split’ between his inward and outward orientations towards her because his public praise reflects his private ruminations on what is praiseworthy.

It might be noticed that the attention patterns demanded by friendship on the attentional account resemble the doxastic practices of the doxastic account, which involved attending to certain sources of evidence, interpretations, and possibilities. In light of this, it’s important to emphasize the differences between the two accounts. The biggest difference is that, on the doxastic account, but not the attentional account, friendship requires doxastic outcomes as well as certain attention patterns. Another big difference is that the attention

37 For the sake of brevity, I will not explore what the attentional can say about each of the arguments for this general intuition that were considered in section 2; I will note, though, it seems to me that those arguments work just as well, if not better, when “positive beliefs” is replaced with “positive attention.”
patterns prescribed by friendship on the attentional account are not properly characterized as doxastic practice since they are not means to doxastic outcomes, but valuable themselves from the perspective of friendship.

It’s also important to emphasize that directing our attention in the ways characteristic of being a good friend on the attentional account isn’t, by itself, good from the perspective of friendship. If they are to reflect well on us as a friend, such attention patterns must manifest certain desires and values – namely, the desires and values involved in mutual caring: a desire for our friends to flourish and achieve their ends, for them to be healthy and happy, and to engage in meaningful relationships and projects. Bommarito makes a similar point about the attention patterns required my modesty. He argues that directing attention in the ways characteristic of modesty isn’t sufficient for modesty: it must stem from the right reasons – that is, from morally good desires and values, or a lack of bad ones. If Jimi Hendrix doesn’t attend to his musical talent only because he’s a pessimist and doesn’t attend to anybody’s good qualities, then his modest-like attention patterns do not count as modesty; if, instead, he doesn’t attend to them because he doesn’t want to ogle at himself, then his attention patterns do count as modesty. Similarly, if Ted directs his attention toward charitable interpretations of Emma’s behavior just because it would feel awkward to admit that his friend acted rudely, then Ted’s pattern of attention doesn’t reflect well on him as a friend. If, instead, he attends to such interpretations because he values Emma’s welfare and wants to defend her reputation, then Ted’s attention direction is good from the perspective of friendship.

This point highlights an important aspect of friendship and its demands: meeting a demand of friendship is good from the perspective of friendship only if it manifests the desires and values at the heart of friendship. We want our friends to meet the demands of friendship, not because of ulterior motives and not because meeting them coincidentally lines up with what they were going to do anyway. We want our friends to meet the demands of friendship because they care about us: it’s part of what being a good friend is.

---

38 Bommarito (2013), p. 103
Weathering the Worries: Developing the Attentional Account

As it stands, there are two competing views about what friendship demands of us cognitively: the doxastic account and the attentional account. The doxastic account entails a surprising claim about epistemology – that, sometimes, our beliefs should be based on things that don’t bear on their truth – and also faces a number of other worries that pressure us to give it up. The attentional account says nothing about belief and so is free from the doxastic account’s controversial epistemological commitments. I now want to consider how the attentional account fares in light of the other worries that trouble the doxastic account and argue it can overcome them. In the process, I’ll flesh out the attentional account a bit; with some meat on its bones, my hope is that it will be more compelling.

The first worry for the doxastic account is that it poses a tension between epistemic rationality and friendship. There is something uncomfortable about saying that an epistemically rational friend is a worse friend and it’s an advantage of the attentional account that its defenders don’t have to say that it is. On the attentional account, friendship doesn’t require us to have certain beliefs, irrational or otherwise. It’s true that, by directing our attention in the ways that friendship demands – generally, towards the positive – we may be more susceptible to forming irrational beliefs. For example, by dwelling on the admirable parts of Rebecca’s poetry, Eric is more likely to be irrationally highly confident that Rebecca’s poetry made a favorable impression on the literary critic. But there’s an important distinction between friendship demanding irrational beliefs and friendship demanding actions that might lead to irrational beliefs. On the attentional account, being a good friend might lead to epistemic irrationality but it doesn’t require it. After all, it’s possible to direct our attention in the way that friendship demands without being epistemically irrational: Eric, for example, can have rational beliefs about Rebecca’s poetry even while dwelling on its good parts.

This feature of the attentional account – that, according to it, friendship can cause, but doesn’t demand, irrationally positive beliefs – is one of its virtues. It’s a fact that good friends tend to have irrationally positive beliefs about each other. But, as we’ve seen, there are problems with claiming along with the defenders of the doxastic account that friendship demands such beliefs. Nevertheless, they seem to be more than merely
The attentional account helps us draw an interesting connection between irrationally positive beliefs and good friendship without making the problematic claim. Since attending in the ways characteristic of friendship can lead to such beliefs, we can say that – and flesh out why – irrationally positive beliefs indicate good friendship, just as a parent’s excessive worry indicates that she is a good parent. Even so, it’s just an indicator. We don’t have to criticize the epistemically rational friend as a worse friend on the attentional account.

Yet it may seem to some that this virtue of the attentional account is masking a vice: if satisfying the attentional demands of friendship can lead us to have biased beliefs or biased evidence about our friends, then the attentional account fails to elude worries about conflicting epistemic and friendship considerations. By directing our attention in the ways that friendship demands, the worry goes, we’re being epistemically irresponsible – we are acting in ways that significantly increase the likelihood of irrational or otherwise epistemically bad beliefs. In this way, it may be thought, the friend who attends disproportionately to her friend’s good side is just as epistemically condemnable as the person who exclusively watches Fox News.

This worry compels us to take a closer look at epistemic irresponsibility. If all actions which significantly increase the likelihood of epistemically bad beliefs are epistemically irresponsible actions, then the attentional account would require epistemic irresponsibility. But I want to suggest that this account of epistemically irresponsible actions is too broad. To modify an example from Horowitz (2019), having low blood sugar makes it significantly more likely that I will form epistemically bad beliefs, but surely I don’t have an epistemic duty to eat a sandwich before reasoning, nor would it be epistemically irresponsible for me refuse to eat one. As Horowitz puts it, “Epistemology should not tell me to have a sandwich.” So there must be some constraint on which actions that increase the likelihood of epistemically bad beliefs can count as epistemically irresponsible. One plausible constraint is the purpose of action: only actions whose primary purpose is doxastic may count as epistemically irresponsible – that is, things we do in order to attain and inform beliefs. If something along these lines is right, then exclusively watching Fox News in order to learn about and understand the things going on

---

40 Horowitz (2019), p. 116
in the world may very well be epistemically irresponsible. But attending to certain features of our friends and ignoring others as friendship demands on the attentional account would not be epistemically irresponsible, even if, like skipping lunch, it increases the likelihood of epistemically bad beliefs. That’s because the purpose of attending in these ways is not attain or inform beliefs about our friends – attention patterns required by friendship on the attentional account are not doxastic practices.

Now for the second worry: on the doxastic account, there are ‘oughts’ of friendship that we cannot satisfy. The attentional account does not face the same worry since we do typically have direct voluntary control over where we direct our attention. For example, a coffee shop patron can choose to dwell on the conversation at a neighboring table instead of the book in front of her; a traveler can choose to focus on capturing a pretty picture for social media instead of contemplating the beauty surrounding her; and Ted can choose to attend to Emma’s endearing passion for the writings of Dostoevsky instead of the report that she acted rudely. On the attentional account, then, the cognitive ‘oughts’ of friendship – namely, directing our attention in certain ways – are ones that we typically can satisfy.

But, of course, there are some cases where we do not have voluntary control over our attention – for example, we cannot turn our attention away from loud sounds or severe pain. On the attentional account, if ought implies can, friends in these cases do not have an obligation to attend in the ways that friendship would require under normal circumstances. If Eric has a migraine during Rebecca’s poetry reading, for example, he does not have an obligation to attend to the poem’s admirable parts. Importantly, though, because we typically do have voluntary control over our attention, “ought implies can” does not render the attentional account incorrect or insignificant. An account that says that friendship often requires friends to φ can still be correct even if there are a few cases where friends do not, in fact, have such an obligation; and an account that says that there are cognitive oughts of friendship in most cases is significant even if there are some atypical cases where, given extenuating circumstances, there are no such oughts.

The final worry is that the doxastic account can’t easily accommodate the idea that friendship requires both impartial thinking and positive thinking. The attentional account can. On the attentional account, to think positively about our friends is to direct attention toward the positive features of our friends and away from the
negative: toward their virtues and away from their vices, toward the impressive features of their work, towards the possibility that they will succeed. To think impartially is to attend to whatever comes to mind – to whatever is salient or striking or simply there to be noticed – whether it be positive, neutral or negative: toward our friends’ virtues when they are acting virtuously and toward their vices when they are acting viciously, toward the unimpressive features of their work as well as the impressive, toward the possibility that they will fail. If friendship requires impartial thinking some of the time in addition to positive thinking most of the time, on the attentional account, this would mean we would have to direct our attention toward the positive most of the time and toward whatever is salient or striking or simply there some of the time.

The attentional account can handle such a requirement because attention, unlike belief, is largely flexible. Switching the focus of our attention from case to case is typically entirely unproblematic, unlike the switching of beliefs. We can direct our attention away from the positive toward the negative and then back again in order to meet various attentional demands – in one case directing our attention toward our friend’s laziness and the next away for it; in one case attending to the possibility that our friend will get into her dream law school and the next attending to the equal possibility that she won’t; in one case dwelling on the impressive features of our friend’s painting and the next on its ordinary features.

6 Friendship and Positive Thinking

Interestingly, it seems that defenders of neither the doxastic nor attentional account can offer general or concise instructions on what to believe or how to direct attention, respectively, in order to be a good friend. But the assumption has been that – whether the cognitive demands of friendship are cashed out in terms of belief or attention – positivity is the common thread running through the variety of thoughts required by friendship. In other words, so far it’s been assumed that friendship generally requires positive thinking, even if it requires impartial thinking in some cases. Now I want to question this assumption: friendship does require positive thinking to some extent, I want to argue, but not the extent that has been assumed. When we step back and

41 I will not venture to account for what determines, with regard to our attention, which things are salient or striking or there to be noticed, though I imagine considerations about morality, prudence, and human psychology all have a say.
look at the heart of friendship, seeking an explanation of why we should think about our friends in certain ways and not others, friendship doesn’t seem to require constant positive thinking with a few exceptions. Rather, it seems to require a significant amount of impartial thinking as well as positive thinking. And if that’s right, I’ll argue, we have even more reason to favor the attentional account – reason that flows from the very heart of what it is to be a good friend.

So why might it be thought that friendship demands positive thinking in most cases? It’s helpful, first, to remember the distinction between fundamental demands and derivative demands and note again that the cognitive demands of friendship are derivative demands – they tell us how to think in order to deepen our care for our friends, or to enable us to advance their welfare and help them to achieve their important ends. Let’s just say that the cognitive demands of friendship tell us how to think about a friend in order to promote the flourishing of the friendship for short. The assumption that positive thinking is the sort of thinking generally demanded by friendship, then, rests on the claim that positive thinking generally has effects both on the friend who is doing the thinking (the subject of thought) and on the friend who is being thought of (the object of thought) that promote the flourishing of their friendship.

Let’s start with the claim about positive thinking and the friend who is the subject of thought. We can look to Stroud for a defense of this claim: given that friendships are partly contingent on esteem for our friends, it is no surprise, she argues, that friendship tells us to “massage our beliefs about our friend’s character in a positive direction” since such positivity promotes esteem. Stroud is getting at something important about human psychology: positive thinking tends to foster attitudes in the subject of thought – esteem, affection, and loyalty, for example – that are good from the perspective of friendship. There is something to the idea that Jill is more likely to like and love Jack, for example, if her thoughts about him emphasize his virtues and deemphasize his vices. These attitudes are good for friends to have towards each other not only because they tend to strengthen the relevant sort of caring in the subject of thought but also because they can motivate her to take steps – and, maybe, giant, sacrificial leaps – to advance her friend’s welfare and help him achieve his ends.

---

42 Stroud (2006), p. 511
I agree that positive thinking about our friends does tend to foster these attitudes to some extent, and, to that extent, friendship requires positive thinking. But it also seems that, for many of us, positive thinking isn’t the only sort of thinking that does this. If — for whatever reason — I want to increase my love and appreciation of my favorite book, yes, I’ll read the best passages over and again but I’ll also reread the whole book. For me at least, and I expect for many others, a balanced mix of positively thinking about how great my friends are and impartially thinking about how they fully are fosters the relevant attitudes best and to the greatest degree. By “thinking about how they fully are,” I have in mind the sort of probing and clear-eyed thinking aimed at trying to understand, to see, to fully take in all of our friend — the whole, messy, beautiful, fascinating person: glaring vices as well as subtle virtues, weakness and the strengths that they belie, quirks and favorite flavors and deepest desires and how they all fit together.

Now let’s consider the friend who is the object of positive thinking. It might be argued that positive thinking about our friends is the sort of thinking that tends to best promote their welfare and help them achieve their ends. This line of argument can be fleshed out by appealing to Keller’s comparison between the benefits of having positively-thinking friends and the benefits of having an encouraging coach. Thinking positively of our friends can motivate and embolden them, which “can make [them] more positive about their prospects, more likely to work hard and hence more likely to improve.” Also, “it can…just make life more pleasant.” In addition to inspiring them to do what it takes to achieve their ends, the argument might go, thinking positively of our friends can boost their confidence and make them happy, thus promoting their general welfare.

Even so, it’s not clear that positive thinking is generally the sort of thinking that best promotes our friends’ welfare and helps achieve their ends; this is largely a contingent matter which depends on individual friends’ unique constellation of desires, goals, and personality traits. Different friends will need different kinds of thinking from us, and the same friend will need different kinds of thinking from us at different times, if we are to best promote their welfare and help them achieve their ends. Perhaps some friends may need positive thinking to significantly outweigh impartial thinking, or even vice versa, but, plausibly, many of us will need significant amounts of both positive thinking and impartial thinking from our friends. Many of us will feel

happy and heartened when our friends think positively of us some of the time but such positivity may seem
disingenuous, cloying, or even patronizing if it’s not balanced out by impartial thinking. Jack may need Jill to
think positively of him on Sunday in order to be encouraging, and, on Monday, he may need her to think
without bias in order to dispense some tough love; I may need you to think impartially about my paper before
I submit it in order to catch errors and then think positively about it after its rejected when I’m doubting that
it’s good enough to submit somewhere else.

To sum up so far: the common thread running through the variety of beliefs, on the doxastic account,
and the variety of attention patterns, on the attentional account, is that they promote the flourishing of the
friendship by affecting the friend who is the subject of thought and the friend who is the object of thought in
ways that are good from the perspective of friendship. But the kind of thinking that does this is not, by and
large, positive. Thus, the common thread in question can’t be simplified to positivity: impartial thinking plays a
much bigger role in producing the relevant effects than has been assumed. It seems, then, that friendship
demands significant amounts of both positive and impartial thinking. And if this is the case, we should favor
the attentional account. To see why, we can appeal to points already made in the discussions above about the
third worry.

Suppose that friendship requires, for example, equal amounts of positive and impartial thinking. On the
doxastic account, this requirement would mean that, half of the time, friendship requires (sometimes irrational)
positive beliefs about the very things that friendship requires rational beliefs about the other half of the time.
As we’ve seen, the doxastic account falters when friendship requires just some cases of rational beliefs in
addition to positive beliefs – we end up with problematic doxastic switacharoos or gross doxastic
inconsistencies. Problems for the doxastic account can only get worse if friendship requires significant amounts
of both positive and impartial thinking. But things look better on the attentional account. On the attentional
account, this requirement would mean we would have to direct our attention toward the positive half the time
and toward whatever is salient or striking or simply there to be noticed the other half. The attentional account
can handle such a requirement because, as we’ve seen, attention is flexible - we can typically easily shift our
attention between the positive, negative, and neutral features of our friends, and it’s no less easy if friendship
demands that we attend not mostly to the positive and occasionally to the negative and neutral, but equally to
them all.

8 Good Friend: Bad Believer?

We’ve seen how the attentional account can overcome the worries that trouble the doxastic account and how
it can accommodate a picture of friendship that requires significant amounts of both positive and impartial
thinking. But there is more that can be said in favor of the attentional account. Let’s conclude by considering a
couple of features of friendship that dovetail especially well with the attentional account.

One feature of friendship that fits neatly within the attentional account is that friendships change over
time: people become better and worse friends as life goes on. Instead of saying that, as the nature of a friendship
changes, so do the friends’ beliefs about each other, the defender of the attentional account can say something
a bit more realistic: as a friendship fades over time, for example, the friends simply attend to each other less
and differently from how they did before. Suppose that a friend of mine, a chef, moves to open a restaurant in
a faraway state, and I have largely lost touch with him and no longer count him as a friend. It would be odd to
say that, when we were good friends, I believed that his cooking rivaled Julia Child’s, but now that I haven’t
talked to him in a year, I believe his cooking is rather run-of-the-mill – after all, my evidence about his cooking
hasn’t changed. More likely, I simply attend to my former friend and his cooking less and differently.

Another feature of friendship that the defender of the attentional account can easily capture is what goes
wrong in cases like the one mentioned earlier in which Jill lands her dream job and, though Jack is outwardly
congratulatory, he is inwardly jealous. Another relevant case might involve long-distance friends, one of whom
always talks about himself during their regular phone catch-ups, and rarely asks about the other. In these cases,
there is something problematic about the way that the friends at fault are thinking about the other, but the
problem isn’t rooted in their beliefs. Suppose that Jack doesn’t believe that it would have been better had Jill not
landed the big job, even though he wished she didn’t. And suppose the self-focused friend doesn’t actually
believe that the goings-on in his life are more important than the goings-on in his friend’s life. The problem,
instead, seems to be that the ways these friends are directing their attention are selfish. The friends at fault are
attending to themselves when it’s appropriate to attend to their friend. What’s more, the ways that these friends
are directing their attention can easily foster attitudes – jealously, pride, self-centeredness – that are inimical to the flourishing of their friendships, which is why attending in these ways seems especially bad from the standpoint of friendship.

In sum: although the recent literature on what friendship requires of us cognitively has focused on beliefs, an attention-based account is better. The attentional account not only accommodates the considerations that motivate the doxastic account and avoids the surprising epistemological commitments and worries that trouble it, but it also affords us the cognitive flexibility that we need to promote the flourishing of our friendships in the ways that we think about our friends. In addition, it equips us with a realistic picture of how our outlook on our friends changes over time and of what goes wrong cognitively in cases where a friend is being selfish in thought. It seems to me, then, that we don’t have to choose between being good friends or good believers: we get to be both.


Chapter 2
The Promising Puzzle

Consider the following three claims.\(^4\)

1. The Evidence Claim: it is rational for you to believe that \(p\) only if \(p\) is supported by your evidence.
2. The Sincerity Claim: your promise to \(\varphi\) is sincere only if you believe that you will \(\varphi\) given that you promise to \(\varphi\).\(^5\)
3. The Permissibility Claim: it is permissible for you to promise to \(\varphi\) only if your promise is sincere and the belief required for sincerity is rational.

On the face of it, each of these claims is plausible. Here’s what’s interesting: if all three are true, it seems that we shouldn’t make many of our promises - specifically, promises to do things which many others, or we ourselves, have set out and failed to do. But surely that can’t be right. After all, those promises are some of our most important! A pair of examples will help bring this issue into focus.

**Wedding** Walt is at his wedding. He is about to promise his partner Wendy to spend the rest of his life with her, for better, for worse. Walt has a lot of evidence that marrying Wendy is the best thing to do: he loves her, and she loves him, and they both very much want to be married to one another. They’ve been together for a while – through ups and downs – and neither has cold feet. But Walt also has evidence that there’s a significant chance that he will not follow through with the promise if he makes it. After all, divorce rates are high, and many apparently equally committed couples have failed to follow through on their vows.\(^6\)

**Presentation** Penny and her colleague Pablo are scheduled to make a joint presentation tomorrow morning at work. Pablo asks Penny to promise him that she’ll be on time. Penny has a lot of evidence that being on time is the best thing to do: her boss would frown upon her and Pablo for starting their presentation late, Penny gets frazzled when she’s running behind and she presents poorly when frazzled, and her tardiness would cause Pablo a lot of undue stress. But Penny also has evidence that there’s a significant chance that she will not follow through with the promise if she makes it. After all, Penny is perpetually late, even to big events – like tomorrow’s presentation – when it’s important for her to be on time.

It seems that Walt and Penny should make their respective promises. In fact, it seems to be not just permissible, but also the best thing for them to do: Walt and Penny want to do that which they would be promising to do, and they are volitionally committed to acting as they would promise; they would be making their promises

---

\(^4\) These claims are modified from p. 22 of Marušić (2015).

\(^5\) The qualification *given that you promise to \(\varphi\)* is a part of the content of the belief necessary for sincerity because, as Liberman (2019) points out (p. 471), “promising to \(\varphi\) usually makes it more likely that you will \(\varphi\) because we are generally motivated to keep our promises.” Importantly, what matters when making a promise is that you believe that you will \(\varphi\) assuming that you promised to \(\varphi\), even if you wouldn’t believe that you will \(\varphi\) assuming, instead, that you didn’t promise to \(\varphi\). That said, for the sake of simplicity, I sometimes leave the qualification implicit when talking about this belief elsewhere in the paper.

\(^6\) This example is adapted from Marušić (2015), pp. 3-4.
voluntarily and happily; they understand what they would be promising and what it would take for them to follow through; following through on the promises is feasible; making and following through on the promises would be good not only for them but also for their promisees; they and the promises would be welcomed by their promisees.47

But Walt and Penny also have evidence that, nevertheless, there’s a significant chance that they won’t follow through on their promises if they make them. This evidence comes from a statistical bad track record in Wedding and a personal bad track record in Presentation. The statistical bad track record in Wedding poses a problematic symmetry between Walt and others who set out to spend the rest of their lives with their partners, a significant percentage of whom failed to do so. The personal bad track record in Presentation poses a problematic symmetry between this time and other times when Penny set out to be one time, a significant percentage of which were times when she failed to do so. However significant the chance involved in the relevant problematic symmetry, Walt and Penny have evidence that there’s an equally significant chance that they won’t follow through on their respective promises.

If this chance is significant enough, then Walt and Penny’s total evidence wouldn’t support the belief that they will follow through. Of course, it’s hard to pin down exactly how significant a chance needs to be in order to count as significant enough. So, for the sake of our discussion here, let’s just pick a number, say 40 percent - if you have an undefeated piece evidence that there is at least a 40 percent chance that you’ll fail to φ, then your total body of evidence doesn’t support the belief that you will φ.

Suppose, then, that Walt knows that 40 percent of couples who set out to stay with each other for the rest of their lives get a divorce, and that Penny knows that she runs late around 40 percent of the time when she sets out to be on time. Suppose, too, that Walt doesn’t know of anything that distinguishes him from these other spouses-to-be, and that Penny doesn’t know of anything that distinguishes this time from these other times. What should Walt believe about whether he’ll stay forever with Wendy, and Penny about whether she’ll be on time?

47 For discussions about the conditions of genuine promising-making, see Fox and Demarco (1996) and Scanlon (1990). See Liberman (2019) for a related discussion on morally permissible promise-making.
If they believe that they will follow through on their promises, then, according to the evidence claim, their beliefs would be irrational. But if they believe in accordance with their evidence and, thus, don’t believe that they will follow through, then, according to the sincerity claim, their promises would be insincere. So – if both the sincerity and evidence claims are true – and if, according to the permissibility claim, a promise should not be made if it would be insincere or if the belief required for sincerity is irrational, then Walt shouldn’t promise to spend the rest of his life with Wendy and Penny shouldn’t promise to be on time. Call this response – the response that renders the verdict that Walt and Penny (and anybody else) shouldn’t promise against the evidence – the hardline response.48

The problem is that the hardline response seems to give the wrong verdicts here: it seems that Walt and Penny should make their respective promises. This problem becomes more troubling when we step back and see that many other cases fit the same mold as Wedding and Presentation – cases involving promises, perhaps, to quit smoking or to be more attentive to a loved one or to stop procrastinating. These are cases in which it seems that promising is the best thing for the protagonist to do even though the track record of others relevantly like them (as in Wedding) or their own personal track record (as in Presentation) suggests that there’s a significant chance (40 percent or more) that they’ll fail to follow through. Call these puzzling cases.

This leave us with a puzzle. We must either reject one of the three claims that generate the problem. Or, in accepting the hardline response, it seems that we must reject the plausible verdicts in puzzling cases – for example, that Walt should promise to spend that rest of his life with Wendy and that Penny should promise to be on time.

In approaching this puzzle, the most obvious strategy is to solve it by denying one of the three claims that generate it in order to show that it is permissible to promise against the evidence. But this strategy is unsatisfying because, as I will argue below, it is morally impermissible to promise against the evidence. I will then argue for the hardline response by showing that it’s not committed to rendering the wrong verdicts in puzzling cases. The hardliner can say that the protagonists in many puzzling cases should make their promises – not because

48 Two terminological notes: (1) The name “hardline response” comes from the idea that those who endorse this response take a hard line on the three plausible claims. (2) The term “promise against the evidence” is shorthand for making a promise when your total body of evidence doesn’t support the belief that you will follow through with the promise.
it’s permissible to promise against the evidence – but rather because, despite appearances, their total evidence does support the belief that they will φ. In many puzzling cases, the argument goes, protagonists’ evidence that there’s a significant chance that they won’t follow through on a promise to φ – the piece of evidence that precludes their total evidence from supporting the belief that they will φ – is defeated, and so, in these cases, the hardliner can say that the protagonists should promise. In puzzling cases where this evidence is not defeated, the hardliner must say that the protagonists should not promise. But, upon closer inspection, this seems to be the right verdict. So, I conclude, the best way to approach the promising puzzle is to show that promising isn’t so puzzling after all: the conditions that give rise to the purported puzzle do not often arise and, when they do arise, they are not problematic.

1 Promising Against the Evidence

Denying any one of three original claims amounts to accepting that, sometimes, it’s permissible to promise against the evidence. If we deny the evidence claim, then there might be cases where it’s rational for you to believe that you will φ even though your evidence doesn’t support that belief. So, in the puzzling cases, it would be permissible for the protagonists to make their promises even though their beliefs that they will φ aren’t supported by the evidence, as long as their beliefs meet whatever rationality conditions are applicable.

If we deny the sincerity claim, then there might be cases where your promise is sincere even though you don’t believe that you will φ. So, in puzzling cases, it would be permissible for the protagonists to promise even when they don’t believe that they will φ, as long as they have whatever attitude is necessary for sincerity.

And, finally, if we deny the permissibility claim, then there might be cases where you should make a promise even if either the promise wouldn’t be sincere or the belief required for sincerity isn’t rational. So, in puzzling cases, it would be permissible for the protagonists to make their promises even if they don’t believe that they will φ, or even if they do believe that they will φ but their belief isn’t rational.
Berislav Marušić develops a solution to the promising puzzle which rejects the evidence claim, and Alida Liberman, as I interpret her, would reject the permissibility claim in favor of subtly different alternative. Their respective approaches are worth considering at length.

Marušić’s solution starts off with a distinction between two ways that you can settle the question of whether you will φ: a theoretical way and a practical way. In order to settle the question in the theoretical way, you must look to your evidence and engage in theoretical reasoning – you must weigh your evidential reasons bearing on whether it’s true that you will φ. If your evidence suggests that you will φ, then you should believe that you will φ. If it doesn’t, you shouldn’t. The outcome of theoretical reasoning is a prediction about whether you will φ.

In order to settle the question of whether you will φ in the practical way, you must first answer the question of whether you ought to φ by engaging in practical reasoning: you must weigh your practical reasons for and against φing. The outcome of practical reasoning is a decision whether to φ. If your reasons tip in favor of φing, you should decide to φ and if they don’t, you shouldn’t. By deciding whether to φ, you indirectly settle the question of whether you will φ. If you decide to φ, the answer is yes, you will φ; if you decide not to φ, the answer is no, you won’t.

When it comes to promising, Marušić goes on to argue, you should settle the question of whether you will φ the practical way – and not the theoretical way. So, according to Marušić, Walt shouldn’t look at his evidence and predict whether he’ll spend the rest of his life with Wendy. Instead, he should weigh his practical reasons for and against doing so and thereby settle, and thus come to believe, that he will. Similarly, Penny shouldn’t look at her evidence and predict whether she’ll be on time. Instead, she should weigh her practical reasons for and against doing so and thereby settle, and thus come to believe, that she will.

Since you settle the question in a practical way in cases of promises, Marušić argues, your belief about whether you will φ is subject to the standards of practical rationality only; theoretical standards don’t apply. In

---

49 Marušić (2013, 2015)
50 I add “as I interpret her” here, not because of any lack of clarity in Liberman, but because she does not directly address the promising puzzle; instead, she argues for several necessary conditions for permissible promise-making and then considers whether promises made under uncertainty can satisfy those conditions. Because of this, my take on Liberman’s approach to our puzzle involves extrapolation.
puzzling cases, the protagonists’ beliefs that they will φ are rational in light of the applicable practical standards of rationality even though they would be irrational in light of the (inapplicable) theoretical standards. In these cases, Marušič concludes, it’s permissible for the protagonists to promise against the evidence — after all, they have the belief necessary for promissory sincerity and that belief is rational.

Now let’s turn to Liberman. She starts by developing four conditions of permissible promise making, one of which regards the reasonableness of promisers’ beliefs that they will follow through on their promises. This condition — she calls it the realistic self-assessment condition — is in the same vein as the condition articulated in the permissibility claim: “A promise is permissible to make only if the promisor blamelessly believes that she will be able to carry out the promised action.”\footnote{Liberman (2019), p. 473}

Liberman goes on to specify that whether you blamelessly (rationally) believe that you will follow through on your promise depends on whether this belief is reasonable in light of a realistic assessment of “yourself and your abilities.” This self-assessment, she then clarifies, takes into account the various internal factors that determine how resolved you are follow through, such as the strength of your will and your susceptibility to temptation.\footnote{Liberman (2019), p. 481} But it does not take into account the various external factors or “outside sources” beyond your control that may prevent you from φing but that would excuse you from the moral censure that normally accompanies failing to fulfill a promise — for example, a car accident that prevents you from fulfilling your promise to pick up your friend at the airport. It will be helpful here to see Liberman as endorsing a qualified version of the permissibility claim: it is permissible for you to promise to φ only if your promise is sincere and the belief required for sincerity is rational given the subset of your evidence regarding the relevant internal factors.

Endorsing this qualified version of the permissibility claim allows Liberman to say that it is permissible to promise in puzzling cases in which the reason why there’s a 40 chance you won’t follow through has to do to external, rather than internal, factors. This entails that, in some cases, it’s permissible to promise against the evidence — namely, when your belief that you will φ is supported by the subset of your evidence regarding the...

---

\footnote{Liberman (2019), p. 473}
\footnote{Liberman (2019), p. 481}
relevant internal factors, but not supported by your total evidence, including considerations about external factors.

2 A Worry for Promising Against the Evidence

Each solution that denies one of the three claims faces its own individual worries. Instead of discussing them all in turn, let’s focus on one general worry that troubles any such solution: it underrates the moral importance of a promiser’s evidential situation.

To feel the force of this worry, it’s helpful to think about the nature of promissory obligation. A promissory obligation is the moral obligation to φ that you incur when you promise to φ. Consider the expectational account. It says that you’re morally obligated to φ when you promise to φ because promising to φ leads your promisee to form an expectation that you will φ. Breaking a promise is tantamount to deceiving your promisee and, because deceiving is wrong, breaking promises is wrong, too. Another account says that you’re morally obligated to φ when you promise to φ because, by promising to φ, you invite the promisee to trust you to φ. It is wrong to break promises because it amounts to a wrongful breach of trust.

If something along these lines is right, there seems to be a serious problem with promising against the evidence. The problem is that it’s wrong to foster in someone the sort of expectation or trust characteristic of promissory obligation if your evidence doesn’t suggest that you will do what that person is expecting or trusting you to do. If you promise to φ when your evidence suggests there’s a significant chance you won’t φ, you put yourself at a correspondingly significant risk of deceiving your promisee or breaching their trust. If you know there’s a significant chance of committing a moral wrong by doing something – in this case, by promising against the evidence – then that’s surely a weighty moral consideration against doing that thing.

53 Marušić’s solution, for example, rests on a controversial view about the rationality of belief and also entails a surprising claim about epistemology: sometimes our beliefs should be based on considerations that don’t bear on their truth. For more worries and criticisms of Marušić’s view, specifically, see Vavova (2018).

54 For helpful surveys and discussions of accounts promissory obligations, see Downie (1995), Shiffrin (2008), Friedrich and Southward (2011), Owens (2012), and Liberto (2016).

55 Scanlon (1990) and Scanlon (1998), chapter 7

56 See Friedrich and Southward (2011) for the trust-based account. Judith Jarvis Thomson develops a similar reliance-based account according to which that you’re morally obligated to φ when you promise to φ because, by promising to φ, you invite the promisee to rely on you to φ. See pp. 294 – 321 of Thomson (1990).
There’s another related, perhaps even more serious, problem with promising against the evidence. When you promise to φ, you give the person you promise evidence that you will φ. Not only that, but you also invite them to take into account that you will φ when they make decisions and plans. After all, a promise to φ involves more than just the corresponding assertion. If Walt promises Wendy that he’ll spend the rest of his life with her, for example, he’s in effect saying, “Rest assured, Wendy, I will spend the rest of my life with you. You can count on it.” It is in light of this that Wendy will expect (trust) him to do so and it is this expectation (trust) which will then inform her decisions and plans.

But what decisions and plans are best for your promisee to make can depend heavily on whether you will φ. If Walt will spend the rest of his life with Wendy, for example, it’ll be best for Wendy to make one set of plans and decisions: to pass up a promotion at work to spend more time with Walt, to have a child with him, to move to the suburbs. If Walt won’t, however, it may be best for Wendy to make a different set: to take the promotion, to remain childless, to stay in the city she loves. If your promisee’s evidence suggests that you will φ, it may be reasonable for them to make the set of plans and decisions that are best given that you will φ; but if your promisee’s evidence doesn’t suggest this, it may be reasonable for them to make a different set. When you promise against the evidence to φ, you give your promisee evidence that makes it the case that it’s reasonable for them to make the set of decisions and plans that are best given that you will φ, even though your own evidence may suggest that it’s reasonable for them to make the other set.

Promising against the evidence is especially problematic in cases like Wedding, when the resulting expectation (trust) affects important, life-altering decisions and plans of your promisee. There’s something reckless and selfish about Walt promising Wendy that he’ll spend the rest of his life with her – knowing that Wendy will make important decisions and plans, and perhaps great personal sacrifice, in light of his promise – when his own evidence doesn’t suggest that he will.57 But the problem remains in cases when there’s less at

---

57 Considering the various ways that societal structures of work and family disadvantage women – and the disproportionate harms that mothers tend to incur in divorce as a result – sheds even more light on the moral importance of promises, particularly on the promises of husbands-to-be in cases like Wedding that involve heterosexual couples intending to have children. As Saul (2003) explains, society is set up so that the best jobs – secure, high-paying full-time jobs – make demands that are extremely difficult for primary childcare-givers to satisfy. Mothers tend to take on most of a family’s childcare responsibilities, which allows fathers to take advantage of the opportunities of the best jobs but leaves mothers financially dependent on their husbands, since most jobs whose demands they can meet are low-paying and part-
time. So it’s no surprise that divorce tends to leave mothers in dire financial straits. Divorced mothers remain unable to meet the demands of the best jobs since they typically have primary custody. What’s more, they no longer have access to their partner’s income since courts rarely grant alimony and child support is often meager; they also now must bear the burdens, financial and otherwise, of single parenthood. Saul notes that, because of all this, 45 percent of households of divorced mothers are at or below the poverty line. See chapter 1 of Saul (2003), especially pp. 8-11. See Okin (1989) for further discussion.

58 Liberman (2019) makes a similar point on p. 471: “it is morally impermissible to allow or actively lead people to form false beliefs, especially about matters that are important to them or relevant to their interests.” It’s worth noting that she goes on to consider ways in which some promises made under uncertainty – including some promises against the evidence – can avoid being morally impermissible in this way. For the relevant discussion, see Liberman (2019), p. 480-481. Ultimately, I do not think that these considerations absolve the relevant promises against the evidence, but will not go into why here – doing so would require a lengthy discussion on subtly different assumptions Liberman and I make about deception, promises, and promissory obligation that would take us too far afield.

59 On Liberman’s view, that there are a number of reasons why it matters whether it’s internal or external factors that make it the case that there’s a significant chance that a promiser will not follow through. Specifically, Liberman argues that it is difficult for promises made under internal uncertainty, but not promises made under external uncertainty, to fulfill the good faith condition – another of her four necessary conditions for permissible promise-making. See section 9 of Liberman (2019). I think she is probably right about this. So, to clarify, I am not denying that whether a promise is made under so-called internal or external uncertainty can make a difference to its permissibility; I am just denying that it explains why it can be morally permissible to promise against the evidence.
What We Want Out of an Approach

If an approach to the promising puzzle is to avoid this worry, it must be some type of hardline response – that is, it must deny that it is sometimes permissible to promise against the evidence. Of course, hardline responses also face their own worry: they seem to render the wrong verdicts in puzzling cases. A good hardline response, then, must address this worry.

There are two broad ways for a hardline response can do this. It can account for why, despite our initial verdict, protagonists in puzzling cases really shouldn’t promise; or, it can show how the hardliner can consistently say that protagonists in these cases should promise. The second way is better than the first. The promises involved in puzzling cases are some of our most important, woven deeply into our relationships and social institutions – a good hardline response can’t issue a blanket prohibition on making them. This means that a good hardline response must show that, in many puzzling cases, the protagonists’ belief that they will φ is supported by their total body of evidence, despite them having a piece of evidence from either a statistical or personal track record that there’s a significant chance that they won’t.

What else do we want out of a response? In the process of proposing his solution, Marušić levels several criticisms against hardline responses. Since we know we want a hardline response, responding to these criticisms forms the basis of a couple of additional desiderata.

Marušić’s first criticism of hardline responses is that they promote what he calls epistemic evasion. Epistemic evasion happens when you refuse to commit to φing – even when φing is clearly the best thing for you to do – just because your evidential considerations don’t allow you to predict that you will φ. In Marušić’s words:

We use epistemic evasion to refuse to take responsibility for our actions. We refuse to commit to doing something, and we take the mere fact that there is evidence that we would fail to follow through if we decided to do it as an excuse—like someone who refuses to commit to being faithful just because he has evidence that there is a significant chance that he will fail. Then our failure is moral and we are blameworthy. Someone who claims to just be an unfaithful person—pointing contritely, but with quiet pride, to a track record of broken trust—is indeed a bastard!

I assume that, in order to count as a good response, a hardline response needn’t show that the protagonists in all puzzling cases should make their promises – just in many. But the hardliner should be able to draw a principled distinction between puzzling cases in which it is permissible for the protagonists to promise and the ones in which it isn’t.

Marušić (2015), pp. 119-120
Another thing we might want out of a hardline response, then, is that it avoid recommending epistemic evasion, or else it explain why it isn’t morally problematic.

Marušić also criticizes hardline responses because he claims that they often can’t account for the asymmetry between what a promiser should believe and what a detached observer should believe about whether the promiser will φ.62 To motivate the thought that there should be such an asymmetry, Marušić starts off by arguing that whether you should believe that you will φ depends in part on whether φing is up to you. He compares two similar cases of belief. Suppose that you want to complete a 5k and a marathon this summer. The 5k is popular, and there’s a lottery that determines who gets to run it – 40 percent of the people who enter the lottery don’t win. The marathon, on the other hand, admits anybody who wants to run it but it is very difficult - 40 percent of people who are admitted fail to complete it.

Before you know the results of the lottery, it seems that you shouldn’t believe that you will complete the 5k: this is a significant chance that you won’t get in and also whether you do is not up to you. But it seems that it would be permissible for you to believe that you will complete the marathon: there is a significant chance that you won’t complete it but whether you do is up to you. That said, Marušić argues, a detached observer should have the same belief in both cases about whether you will complete the race in question since the chance that you won’t complete it in each case is equally significant.

In all cases of promising, the relevant belief is about whether the promiser will do something that’s up to them. So promisers’ beliefs about whether they will φ should be different from a detached observer’s belief about what they will φ – even if the promiser and detached observer have the same evidence. Walt’s belief about whether he will spend the rest of his life with Wendy should be different from a sociologist’s belief about the same thing, for example, even though both Walt and the sociologist have access to all the same information about divorce rates.

Marušić argues that hardline responses often can’t account for this asymmetry, including in puzzling cases – after all, both the protagonist and a detached observer can have access to the same bad track record which would give them each evidence precluding their respective total bodies of evidence from supporting the belief

---

62 Marušić (2015), pp. 20-22
that the protagonist will $\varphi$. According to the claim about evidence which the hardliner accepts, then, both the protagonist and a detached observer shouldn’t believe that the protagonist will follow through. So something else that we might want out of a hardline response is that it can account for the asymmetry which Marušić says it can’t.

4 The Encroachment Response

We want a hardline response that shows that, in many puzzling cases, the protagonists’ total body of evidence does support the belief that they will $\varphi$. Considering a potential hardline response that doesn’t work – the encroachment response – will help point the way toward one that does.

The encroachment response is based on an idea at the heart of a family of epistemological views according to which the rationality of belief is, in part, determined by non-evidential considerations. These views are compatible with the evidence claim because of the particular role they say non-evidential considerations play: non-evidential considerations set the evidential threshold for the rationality of belief. Your belief that $p$ might still be rational only if it’s supported by your total evidence, the thought goes, but practical or moral considerations determine how much evidence is needed in order for your belief to count as supported.\(^63\) If the stakes for getting it right about $p$ are high, or if believing $p$ comes at significant cost, the evidential threshold is set high. If the stakes for getting it right about $p$ are low, and if believing $p$ comes at little to no cost, the evidential threshold is set relatively low.

It might be thought that the hardliner can appeal to encroachment views to develop a response to the promising puzzle: in puzzling cases, the evidential threshold for the protagonists’ belief that they will $\varphi$ is set by non-evidential considerations. The question arises, what non-evidential considerations are relevant, and how high do they set the evidential threshold? There are two relevant considerations – a practical consideration and

---

a moral consideration – each of which sets the evidential threshold at a different height. Both encroachment solutions based on these considerations fail give us something we want.

The relevant practical consideration is that making the promise in question is the best thing for the protagonists to do: after all, their practical reasons favor promising. This practical consideration implies that, in puzzling cases, the stakes for getting it right are low, and believing that they will follow through comes at little practical cost and, if anything, practical benefit. What’s important is that the protagonists have the particular belief that they will ϕ so that they can sincerely promise, not that they get it right answer about whether they will ϕ. So the practical consideration sets a relatively low evidential threshold. If the evidential threshold in puzzling cases is low – the thought goes – then maybe protagonists do have enough evidence to support the belief that they will ϕ, which would allow the hardliner to render the right verdict in these cases.

The problem with this response is that it is subject to the same worry that troubles promising against the evidence. Even if Penny, for example, does have enough evidence – because of a low evidential threshold – for her belief that she will be on time to be rational, she doesn’t have enough to avoid taking on a significant risk of deceiving, or breaching the trust of, Pablo. She also shouldn’t invite Pablo to base his plans and decisions on the assumption that she will be on time if she doesn’t have a lot of evidence that she will be - at least enough evidence that would rationalize her belief on a moderate evidential threshold.

This leads us to the encroachment response based on the relevant moral consideration, which is that a lot is hanging on whether the protagonists will follow through on their promise. The moral consideration raises the stakes for getting it right about whether they will ϕ, thus setting a high evidential threshold for the corresponding belief to be rational.

But this is no response at all. If the evidential threshold is high, it’s even clearer that the protagonists’ belief that they will ϕ is not rational. The encroachment response would entail that they should not make their promises. Not only that, but if the evidential threshold is high, the verdicts become even more implausible. Some beliefs that were rational on a moderate evidential threshold are no longer rational on a high evidential threshold meaning that there would be even more puzzling cases. In all of these cases, the encroachment response renders the wrong verdict.
5 The Hardline Response

Our assumption from the outset has been that if you have a piece of evidence that there’s a 40 percent chance that you won’t φ, then your total body of evidence doesn’t support the belief that you will φ. But, of course, the assumption holds only if this piece of evidence is undefeated. The hardline response I want to develop – which I will refer to simply as the hardline response – centers on the claim that, for protagonists in many puzzling cases, this piece of evidence is defeated.

This hardline response starts off with a natural thought: sometimes, your evidence supports the belief that you will φ even though many others, or you yourself, have set out and failed to φ. It’s easy to come up with examples: Serena Williams’ evidence supports the belief that she will qualify for Wimbledon even though many others before her tried and didn’t qualify; the evidence of a high-polling candidate for congress may support the belief that she will win the election even though many others ran for that seat and lost; a student’s evidence may support the belief that he will pass his math exam on Monday – he’s started seeing a tutor – even though he has failed math exams many times before.

In each of these cases, the person in question has good reason to believe that they are not relevantly like the other people, or their past self at other times, who set out to φ but failed, and that’s why their chance of success cannot be read off the success rate of those other people or times.

The hardline response says that this is what happens in many puzzling cases: the protagonists have good reason to think that they are not relevantly like the comparison group - the other people cited by statistical bad track records or their past self at the other times cited by personal bad track records. This good reason defeats the piece of evidence they get from the bad track record which cites that comparison group. After all, if the protagonists have good reason to believe that they are not relevantly like the comparison group, then it is not reasonable for them to infer that their chance of success mirrors the success rate of the comparison group and, thus, the track record no longer provides them with grounds for believing that there’s a significant chance that they won’t φ.

With that piece of evidence defeated, the protagonists’ total evidence is no longer precluded from supporting the belief that they will φ. If the protagonists’ total evidence now does support that belief, and if the
protagonists so believes, they have the belief that is necessary for promissory sincerity and that belief is rational. In these cases, the protagonists’ promises satisfy the conditions of the permissibility claim and so the hardliner is in a position to render the right verdicts – the protagonists should make their promises. My task in the remainder of the paper is to look more closely at puzzling cases of this sort and, in the process, explain why we should think that many puzzling cases are of this sort and why we should accept the hardliner’s verdicts in the puzzling cases that aren’t.

6 The Hardline Response: No-symmetry Cases

Let’s first look at puzzling cases in which the protagonists have a good reason from the start to believe that they are not relevantly like the comparison group. Call these no-symmetry cases: the protagonists’ evidence suggests that the problematic symmetry posed by the relevant track record doesn’t hold to begin with. In these cases, the protagonists have evidence about themselves or their circumstances that suggests that they are not relevantly like the comparison group. Broadly, there are two main ways this evidential situation can play out.

The first way is that the protagonists have evidence suggesting that they are much more likely than the comparison group to φ. This may be because the protagonists can point to a specific feature that distinguishes them in this way from the comparison group. Suppose that Penny knows that she cares a great deal about being on time tomorrow: in addition to being her colleague, Pablo is a dear friend, and Penny is eager to avoid stressing him out. Or suppose, for another example, that a protagonist has tried and failed to quit smoking many times before but, this time, a baby is on the way.

Or this may be because the protagonists have evidence about themselves or their circumstances which suggests they are very likely to φ, and, from this, they reasonably infer that they are much more likely to φ than the comparison group. Suppose that Walt knows that he and Wendy have the same core values, compatible personalities, and complementary strengths and weakness; he desires being in a healthy marriage over pretty much everything else; he and Wendy have built a stable, loving relationship over the years; both sets of their parents have fostered flourishing marriages; and – to top it off – they belong to a strict religious community whose members rarely ever cheat or divorce. Walt knows that all of this makes it very likely that he will spend
the rest of his life with his partner, and, thus, he reasonably infers that he is much more likely to do so than members of the comparison group, whose average success rate of staying together is mediocre. Note that this inference is reasonable even if Walt does not know much at all about the members of the comparison group besides their average success rate.

The second way that this evidential situation can play out is that the protagonists have evidence suggesting that the features they share with the comparison group are not telling of their chance of success at φing. In order to be relevantly alike, the protagonist and comparison group must share relevant features – namely, features that are the same as, or related to, the features that are causally responsible for their respective chances of success. They are not relevantly alike if the shared features are something irrelevant, like favorite ice cream flavor or hair color. After all, if what’s causally responsible for the comparison group’s success rate is not the same as, or related to, what the group and protagonist have in common, then it would be a mere happy accident if the group had an average success rate that reflects the protagonist’s chance of success.

Suppose that Walt has a decent grasp on the features – the values, interests, desires, goals, personality traits, circumstances, and so on – that go into determining how likely it is that a marriage will last. Suppose, too, that he has evidence – gathered from a lifetime of interacting with a plethora of people in his demographic group – which suggests that couples demographically similar to him and Wendy vary greatly when it comes to those features. Since demographic features are not causally responsible for making a marriage last, nor are they correlated with the features that are, Walt has good reason to believe that he is not relevantly like the comparison group.

To sum up so far: in no-symmetry cases, protagonists have good reason from the outset to think that they are not relevantly like the comparison group. This may be because they have evidence that they are much more likely than the comparison group to φ, or because they have evidence that the features they share with the comparison group are not very telling about their chance of success. In all these cases, the evidence that precludes the protagonists’ total evidence from supporting the belief that they will φ is defeated. Assuming that the protagonists’ total evidence now does support this belief, and that they so believe, the hardliner is in a position to render the right verdict: they should make their promises.
In no-symmetry cases where the total evidence does not now support this belief, the hardliner must say that protagonists shouldn’t promise. But this isn’t a ding against the hardliner. Suppose, for example, that Penny can point to a feature making her much less likely than normal to be on time. Although this is sufficient for a good reason to believe that she is not relevantly like the comparison group, Penny’s evidence does not now suggest that she will be on time. So, in this case, the hardliner must say that Penny should not make her promise. And this seems to be the right verdict.

7 The Hardline Response: Broken Symmetry Cases

Now let’s look at puzzling cases in which the protagonists do not start off with a good reason to believe that they are not relevantly like the comparison group. Even in these cases, their total body of evidence can end up supporting the belief that they will φ if they gain such a reason in the process of deliberating about whether to promise. Call these broken symmetry cases: the protagonists’ evidence initially suggests the problematic symmetry holds but they get a new evidence suggesting it doesn’t.

In some broken symmetry cases, protagonists gain the sort of evidence that they had from the start in no-symmetry cases. For example, suppose that Walt introspects and identifies the desires, values, beliefs, and goals that make him very likely to spend the rest of his life with Wendy only after being confronted with the discouraging statistics about divorce.

But, I want to argue that, in other broken symmetry cases, this evidence comes from a special source – namely, a special sort of seeming or inclination. Imagine a protagonist in a statistical bad track record puzzling case who is deliberating about whether to promise to do something that many others relevantly like him have set out and failed to do. When confronted with the track record, it strikes him that he is more committed or has more determination or wants it more than they do – be is different from them. Or imagine a protagonist in a personal bad track record puzzling case who is deliberating about whether to promise to do something that she, herself, has set out and failed to do many times over. When confronted with the track record, it strikes her that she is more committed or has more determination or wants it more this time – this time is different. These are the special sort of seemings in question. Call them difference-seemings.
A few general points about seemings are in order. Roughly, a seeming is a phenomologically-laden impression that the proposition you are considering is true. The phenomenology characteristic of seemings is striking-as-true. Seemings and beliefs often go together. When you believe, say, that you will complete a marathon, it might also strike you as true that you will complete a marathon. But seemings and beliefs can come apart. You can have a seeming that something is true without the corresponding belief and vice versa. This is important because, in broken-symmetry cases, protagonists have a difference-seeming without the corresponding belief. In fact, the difference-seeming ends up providing the grounds for the corresponding belief.

Seemings come in degrees of strength. In other words, a proposition can strike you more or less as true. Your seeming that 2+2=4, for example, is probably stronger than your seeming (supposing you have one) that your favorite normative ethical theory is correct. This is important because, for a difference-seeming to be worth much, it must be fairly strong. It also must be fairly steady. A difference-seeming won’t count for much if it flickers in and out.

With that, we get to the main contention here: the presence of a strong and steady difference-seeming can sway the permissibility status of a promise in puzzling cases. Consider the following variations on Wedding and Presentation.

**Wedding Variations** Walt knows that spending the rest of his life with Wendy – and promising her to do so – is the best thing for him to do. But he also knows that 40 of couples demographically to him and Wendy wind up divorced. Walt doesn’t have any initial reason to believe that he is relevantly different from demographically similar spouses-to-be, many of whom failed to follow through on their vows. **Variation 1** Nevertheless, it strongly *seems* to Walt that he is more committed – whenever Walt contemplates his situation, it strikes him that he is special, that his level of commitment to Wendy far outruns that of the average millennial. **Variation 2** Moreover, it doesn’t seem to Walt that he’s any more committed – whenever he contemplates his situation, either he doesn’t have a seeming regarding whether he is more committed, or he does have a seeming, but his seeming is that he’s not any more committed.

**Presentation Variations** Penny knows that being on time to the presentation – and promising Pablo to be on time – is the best thing to do. But she also knows that she runs late 40 percent of the time when she sets out to be on time. Penny doesn’t have any initial reason to believe that there’s something about her or her circumstances that makes her more likely than normal to be on time. **Variation 1** Nevertheless, it strongly *seems* to Penny that she is very determined to be on time tomorrow – when Penny contemplates her situation, it strikes her that this time is special, that her level of determination far outpaces her normal level. **Variation 2** Moreover, it doesn’t seem to Penny that this time’s any different – when she contemplates her situation,

---

64 If you don’t like talk of seemings, you can instead think of difference-seemings as difference-inclinations or difference-hunches: the protagonists have an inclination / hunch that it’s true that they’re different or that this time is different.
either she doesn’t have a seeming regarding whether she’s very determined, or she does have a seeming, but
the seeming is that she’s not any more determined than normal.

The main contention, as applied to these cases, is that the protagonists should make their promises in the first
variations, but not in the second variations, and the reason for this difference in verdicts is the presence, or lack
thereof, of a difference-seeming. Difference-seemings make a difference to the rationality of the protagonists’
belief that they will φ: the belief is rational, and therefore the promise is permissible, in the first variations but
not in the second variations.

Of course, this contention is plausible only if difference-seemings are epistemically significant. The idea is
that when the protagonists have a strong and steady difference-seeming, they gain evidence that they are, in fact,
different from the other people cited by a bad statistical track record or that this time is, in fact, different from
the other times cited by a bad personal track record. This evidence gives the protagonists good reason to believe
that they are not relevantly like the comparison group, thus breaking the problematic symmetry posed by the
bad track record.

The idea that seemings are epistemically significant isn’t new. Phenomenal conservatives have been
defending that idea for a long time – and apply epistemic significance to seemings across the board.65 But
phenomenal conservatism is not uncontroversial. It would be better, then, to have an account of the rational
basis of difference-seemings that’s acceptable to those who reject phenomenal conservatism. Such an account
would just need to explain why difference-seemings, in particular, have a rational basis.

Here’s a suggestion: difference-seemings generally arise from internal features that determine the truth of
their propositional content. So, for example, Walt’s and Penny’s difference-seemings arise from affective and
motivational states – the various cares, values, desires, goals, and so on – that go into determining their high
levels of commitment and determination, respectively.

Taking this suggestion on board helps clarify the role difference-seemings play in broken symmetry cases.
For whatever reason, the protagonists in these cases cannot put their finger on the features that give rise to

65 Phenomenal conservatives defend principles like this: if it seems to you that p, then, in the absence of defeaters, you
thereby have at least some positive justified reason for believing that p. For defenses of this principle (or principles close
their difference-seemings. Perhaps they simply haven’t introspected, or perhaps they have introspected but these features are occluded from them – they just can’t get a good read on them or they downright misread them. Their difference-seemings, then, serve as a proxy for knowledge of these features. They indicate to the protagonists that there is something about them that makes them, for example, very committed or very determined, even though they can’t place their finger on what, exactly, it is.

But, notably, the seeming isn’t just they are very committed and very determined – it’s that they are more committed and more determined. Difference-seemings, then, catch an implicit inference from high levels to higher levels – Walt’s inference that his high level of commitment must be higher than the other spouses-to-be cited who have a mediocre success rate of staying together and Penny’s inference that her high level of determination this time must be higher than at the other times cited given her mediocre success rate of being on time.

But why think that this suggestion is correct? Whether difference-seemings do, in fact, generally arise from the relevant truth-related features is an empirical question. Psychological studies would need to test whether promisers who report having a strong and steady difference-seeming before promising are more likely than those who report otherwise to follow through on their promises. If these studies showed that the rate of follow-through was significantly higher among promisers with difference-seemings than without, we would have empirical proof that difference-seemings are generally grounded in that which makes you likely to follow through.

That said, if we assume some sort of internalism about epistemic rationality, it doesn’t have to be the case that difference-seemings do, in fact, generally arise from truth-related features in order for them to be epistemically significant; it just has to be the case that the suggestion that they do is plausible. This is important because it means that we don’t have to wait on empirical work to know whether difference-seemings are epistemically significant. We just have to look and see whether the suggestion is plausible given the evidence currently available. So, is it plausible? I think it is.

It’s helpful here to reflect on some familiar personal experiences. Consider times when it seems to you that you are more or less likely than normal to do something even though you don’t know why. It may seem evident
to you, for example, that you will not get on the Ferris Wheel at the fair with the rest of your friends even though it’s never dawned on your that you’re acrophobic. Or it may strike you as true that you will join in on weekly drinks with your colleagues, which you normally pass on – even though you don’t realize it, you are avoiding the mountain of grading waiting for you at home. Or consider times when you have a seeming about you would do in some hypothetical dangerous scenario. It might seem obvious to you would run away in a frenzy (or that you would confront the danger head-on) even though you aren’t aware that you’re prone to panic (or admirably brave). I take it that such experiences are familiar, which implies that seemings about prospective actions can and do arise from unknown features that help determine the truth of their propositional content. And, if that’s the case, then the suggested explanation of why difference-seemings have a rational basis, far from being ad hoc or new-fangled, seems plausible.

So, we do have some reason in favor of the suggestion. One might worry, though, that existing empirical work on self-serving cognitive biases gives us better reason to think that difference-seemings arise from such biases. The phenomenon of illusory superiority, for example, is well-documented in psychology. Studies show that people typically report that they have positive characteristics and skills to a higher degree, and negative characteristics and ineptitudes to a lower degree, than the average person. The thought, then, is that it’s more likely that difference-seemings arise, not from relevant truth-related features, but rather from self-serving cognitive biases like illusory superiority.

Thankfully, this worry is less troublesome than it may first appear, for a couple of reasons. First, illusory superiority has to do with comparison to others, so it could only give rise to difference-seemings in puzzling cases involving statistical bad track records. Second, the relationship between the relevant empirical work and difference-seemings is too loose to safely infer from this work that even these difference-seemings likely arise from illusory superiority. Puzzling cases involve a seeming that’s triggered when protagonists contemplate a particular action along with how likely they are to do that action at some point in the future. The relevant empirical works shows that people tend to rate themselves better than the average person when faced with a list of amorphous character traits or skills, untethered to any particular action. It would be a far stretch infer

---

66 Alicke (1985) and Hoorens (1993)
from these studies that – in the specific circumstances of puzzling cases – illusory superiority likely causes it to seem to protagonists that they are more committed or determined to $\phi$, for example, than they actually are.

This stretch becomes even further when considering the relevant sense in which protagonists who are more determined or committed to $\phi$ are superior to the comparison group is that they are more likely to $\phi$, but this narrow sort of superiority doesn’t seem to be closely related to superiority of character or skill that illusory superiority tends to distort when you evaluate yourself. It’s doubtful, then, that a tendency to see yourself as superior without warrant when it comes to character traits and skills would cause you to see yourself as superior without warrant in this narrower way related to your likelihood to $\phi$.

Where, then, should we come down on the epistemic significance of difference-seemings? If pheromonal conservatism is true, then difference-seemings – like all seemings – are epistemically significant. But, if pheromonal conservatism turns out to be incorrect, difference-seemings are still epistemically significant just in case the suggestion that they arise from truth-relevant features is plausible. Even though we lack empirical proof for this suggestion, reflecting on familiar personal experiences lends does lend this suggestion some plausibility, and existing psychological studies on self-serving cognitive biases do not undermine it. It seems, then, that there’s some – though defeasible and not decisive – reason to accept difference-seemings as epistemically significant.

In sum: in broken symmetry cases, the protagonists do not start off with good reason to believe that they are not relevantly like the comparison group, but they gain it in the process of deliberating whether to promise. In some broken symmetry cases, it’s been argued, the evidence that gives them this good reason is provided by a difference-seeming. Assuming that difference-seemings are epistemically significant, they give protagonists

---

67 It’s true that having certain character virtues to a high degree – having a lot of grit, for example, or being very trustworthy – may make it more likely for you to follow through on any given promise. But, generally, your level of commitment or determination to $\phi$ depends heavily on value-neutral features of your personality and circumstances – your particular preferences, priorities, and goals. This is why it’s doubtful that a (perceived) high level of determination or commitment to $\phi$ is reliably correlated with (perceived) superiority in character or skill.

68 I’ve focused on the limits of empirical studies on illusory superiority since it seems to be especially pertinent to this discussion. That said, I assume that similar things can be said about the limits of empirical studies on other self-serving cognitive biases.

69 In section 9, there is a discussion about what it means for the hardline response if we learn that difference-seemings do not, in fact, arise in the suggested way, in which case there would no longer be reason to accept them as epistemically significant.
evidence that they are more likely to φ than the other people cited by a statistical bad track record or their past self at the other times cited by personal bad track records. This evidence provided by difference-seemings defeats the evidence provided by the track record that there’s a significant chance they won’t φ, thus allowing the protagonists’ total evidence to support the belief that they will φ.

8 What About Untrustworthy Seemings?

It might be worried that protagonists’ difference-seemings can arise from wishful thinking or self-deception, in which case it would not rationalize their belief that they will φ. We might think of friends who consistently promise and fail to return borrowed books or pay us back for dinner, and also of more serious cases – an absent father, for example, who consistently fails to follow through on promises to show up to his child’s parties and plays and recitals.

It seems that the hardliner is committed to saying that, if these people have a difference-seeming, their beliefs that they will φ is rational. This would be a troubling commitment, especially if we know that the absent father, for example, has a history of untrustworthy difference-seemings when he contemplates whether he will show up – surely it’s not rational for him to believe that he will this time.

But the hardliner isn’t so committed. That’s because the evidence that difference-seemings provide can, itself, be defeated. Suppose, for example, that Penny regularly has difference-seemings when she contemplates whether she’ll be on time yet still fails to be on time. Even when it seemed to her in the past that this time is different, she continually proceeded to run late. In this case, Penny has evidence that her difference-seeming does not indicate that she is, in fact, more likely than normal to be on time. The evidence that gives her reason to believe that she is not relevantly like the comparison group provided by such a difference-seeming, then, would be defeated.

Of course, the evidence provided by difference-seemings can be defeated by things other than a history of untrustworthy ones. Suppose that, in addition to his strong and steady difference-seeming, Walt knows something about himself that makes him very unlikely to spend the rest of his life Wendy – say he knows that he as a strong distaste for monogamy that Wendy doesn’t share. In this case, his evidence of his attitude towards
monogamy defeats the evidence from the difference-seeming that he is more committed than the average spouse-to-be.

In broken symmetry cases where the evidence provided by a difference-seeming is defeated, it is not rational for protagonists to believe that they will \( \varphi \) - even if they have the seeming in full force. If the evidence provided by the difference-seeming is defeated, it doesn’t serve its own defeating purpose – it doesn’t defeat the evidence from the relevant bad track record that there’s a significant chance that they won’t \( \varphi \). So, in these cases, the protagonists’ total evidence remains unable to support the belief that they will \( \varphi \) and, thus, the hardliner must say that they shouldn’t make their promises.

But this seems to be just the right verdict. It seems that Penny should not promise, and neither should Walt. More generally, it isn’t permissible for any protagonist with a history of untrustworthy difference-seemings – the book-borrower, the free-riding dinner companion, or the absent father – to make promises based on another difference-seeming.

9 Evaluating the Hardline Response

Let’s conclude by looking at whether the hardline response gives us what we want out of an approach to the promising puzzle. One thing we want is that it avoid the general worry troubling solutions that entail that it is sometimes permissible to promise against the evidence. The worry was that, by promising against the evidence, you not only take on a high risk of deceiving, or breaching the trust of, your promisee but also you invite your promisee to make decisions and plans based on the assumption that you will follow through, even though your own evidence doesn’t suggest that you will. The hardline response avoids this worry because it maintains without concession the impermissibility of promising against the evidence: if your total body evidence does not support the belief that you will \( \varphi \), you should not promise to \( \varphi \).

Another thing we want out of a hardline response is that it render the right verdicts in puzzling cases. Puzzling cases are those in which it seems that the protagonists should promise against the evidence. The hardline response shows that, despite appearances, in many puzzling cases, the protagonists’ total evidence does support the belief that they will \( \varphi \), and so the hardliner can say that they should promise. The hardliner renders the right verdicts in these cases.
Now, it’s true that there are some puzzling cases where the hardliner cannot say that protagonists should promise. These include puzzling cases like the second variations of Wedding and Presentation – cases in which protagonists do not have good reason from the start to believe that they are not relevantly like the comparison group, nor do they gain it. But the hardliner gets it right here: it would be impermissible for Walt and Penny to promise in the second variations of their respective cases.

The hardliner must say that protagonists shouldn’t promise in a few more types of puzzling cases, too – namely, in no-symmetry cases where protagonists have good reason to think that they are not relevantly like the comparison group but in ways that make them less likely to φ, and also in broken symmetry cases where the evidence provided by the difference-seeming is defeated. But, as we’ve seen, this seems to be the right thing to say. So, both in puzzling cases where protagonists should promise and also in puzzling cases where, upon closer inspection, it’s clear they shouldn’t, the hardline response renders the right verdicts.

It’s also worth addressing what happens to the hardline response if psychology shows us one day that difference-seemings do not generally arise from features that help determine the truth of their content, in which case the suggestion that they do loses its plausibility and difference-seemings would no longer be epistemically significant. The downside would be that, on the hardline response, there would be considerably fewer cases in which promising is permissible. Perhaps this would make the hardline response a bit less satisfying, but it’s worth emphasizing that the hardliner could still say that the protagonist should promise in a large number of puzzling cases – namely, no-symmetry cases and broken-symmetry cases where the relevant evidence is provided by something other than a difference-seeming.

More importantly, the hardliner seems give the right verdict here, however unsatisfying. In learning that difference-seemings do not arise from truth-related features, we would learn something morally important, albeit unhappy, about ourselves that we should take into account when making promises for the sake of our promisees: in light of this new self-understanding, it would be morally impermissible for protagonists with difference-seemings to make their promises. The hardliner, then, gets it right both on the assumption that difference-seemings are epistemically significant and on the assumption that they’re not.
Yet another thing we want out of a hardline response is that it avoid recommending epistemic evasion, or else explain why it’s not morally problematic. The hardline response can. First, it’s worth noting that the potency of the worry that hardline responses promote epistemic evasion is diluted quite a bit – as we’ve seen, the hardliner can say that protagonists should promise in many puzzling cases. This means that the hardliner recommends the line of reasoning characteristic of epistemic evasion – “well, since I have evidence that there’s a significant chance I won’t follow through, then I just won’t make the promise” – a lot less than Marušić contends.

And in the cases where the hardliner does recommend this line of reasoning – namely, in the puzzling cases just surveyed where the hardliner must say that protagonist shouldn’t promise – it’s exactly the line to take. When protagonists refuse to commit and promise to $\phi$ in these cases, they are not shirking responsibility for their future actions. Instead, constrained by self-awareness and caution, they’re taking steps to lessen the damage of their future actions. That’s not evading responsibility – that’s taking it!

The hardline response can also account for the asymmetry between what the protagonist should believe and what a detached observer should believe about whether the protagonist will $\phi$ in puzzling cases – at least in those puzzling cases where the protagonists should promise. In these cases, it turns out, the protagonists’ total body of evidence supports the belief in question while the detached observer’s total body of evidence doesn’t. That’s because the protagonists – but not the detached observer – have access to the evidence that provides them with good reason to believe they are not relevantly like the comparison group.

In no-symmetry cases, this evidence largely comes from knowledge of the various desires, cares, values, beliefs, and other motivational and affective states that determine whether the protagonists will $\phi$. Many such states are mental states, and, of course, protagonists have unique access to their own mental states. So protagonists alone – not the detached observer – have unique access to the evidence about these features which provides them with good reason to believe that they’re not relevantly like the comparison group.

In a good many broken symmetry cases, I’ve suggested this evidence comes from a difference-seeming, which is also a mental state, so the same thing can be said. What’s more, only the possessor of a difference-
seeming can acquire good reason to believe its propositional content – since a difference-seeming is had only by the protagonist, not by detached observers, only the protagonist reaps its epistemic benefits.

To conclude: the hardline response gives us everything we want out of an approach to the promising puzzle. Not only does it render the right verdicts in puzzling cases and concedes no ground on promising against the evidence, but it also avoid recommending epistemic evasion and makes room for the doxastic asymmetry between promiser and detached observer. This means we can keep our three original claims and also our promises to do things that many others, or we ourselves, have tried and failed to do, all without bearing the moral burdens of promising against the evidence.
Works Cited


Chapter 3
Prejudiced Beliefs Based on the Evidence
Responding to a Challenge for Evidentialism

It is widely assumed that prejudice – racism, sexism, anti-Semitism, homophobia, and the like – necessarily involves some sort of epistemic flaw: prejudice misconstrues reality, it neglects relevant information, it misleads and misinterprets and invites fallacious reasoning. Correspondingly, prejudiced beliefs are widely assumed to be epistemically irrational. This assumption will be our starting point.

According to a common view about epistemic rationality, only evidential considerations get a say in what is epistemically rational to believe: S’s belief is rational iff S’s evidence supports its propositional content and it is properly based on that evidence. Call this view evidentialism about epistemic rationality (evidentialism for short).70

Since prejudiced beliefs are epistemically irrational, evidentialists have to say that prejudiced beliefs must not be properly based on evidence that supports them. Paradigmatic prejudiced beliefs fit nicely within an evidentialist framework. Consider, for example, the anti-Semite’s belief that Jews are conspiring for world domination or the sexist’s belief that Jack must be smarter than Jill. Beliefs like these go against the evidence and so evidentialists can accurately categorize them as irrational and prejudiced.

But recently, philosophers have been interested in cases of belief that seem to undermine evidentialism. These are cases of belief that seem to be both prejudiced and properly based on the evidence. In these cases, believers have strong statistical evidence that members of a certain social group have some property and then come to believe that an individual member of that social group will likely have that property. Here are is one such case.

---

70 A few terminological notes are in order. “Evidentialism” is used to pick out several different views in contemporary epistemology. Throughout the paper, “evidentialism” exclusively picks out this account of epistemic rationality. Importantly, this account concerns doxastic rationality, rather than propositional rationality. I will use “rationality” to refer to epistemic rationality of this sort from here on. “Evidence” is meant, broadly, to include all truth-related considerations. A belief “based on the evidence” is sometimes used as shorthand for a belief that’s properly based on evidence that supports it.
The Server  Spencer works as a server at a restaurant. He sensed that White diners tipped more than Black diners. He did a bit of research online, and found a well-documented social trend that Black diners tip substantially below average. Spencer weighs the evidence before reaching his belief that Black diners tip substantially below average. A Black diner, Jamal, enters Spencer’s restaurant and dines in a booth outside of Spencer’s section. Spencer believes that Jamal will likely tip below average.71

There are many cases like The Server. Imagine a business consultant believing that the woman he sees walking out of a corner office is likely an administrative assistant given evidence that the vast majority of women who work at that office are administrative assistants. Or imagine a police officer believing that a Hispanic male from an inner-city neighborhood is likely a gang member given evidence that most Hispanic males from that neighborhood belong to gangs.72

Some philosophers appeal to cases like The Server in order to challenge evidentialism.73 This challenge starts with two intuitions about beliefs like Spencer’s. The first intuition is that Spencer’s belief is racist, which entails that it’s irrational. The second intuition is that Spencer’s belief is properly based on his evidence. If both of these intuitions are veridical and Spencer’s belief is irrational but properly based on his evidence, then evidentialism is false.

To respond to this challenge, evidentialists must account for these intuitions. In other words, evidentialists need to explain why beliefs like Spencer’s seem to be both prejudiced and properly based evidence that supports it if they, in fact, are not. My goal in this paper is to develop such a response by offering an account of mistaken intuitions.

Moving forward, it’s important to bear in mind that cases like The Server are not merely interesting epistemologically, but significant morally: they center on a type of hurtful experience had by real people, and relate to number of moral issues such as racial profiling and implicit bias that bring real harm. So, although the main focus here is on the epistemic dimension of these cases, extra care and sensitivity to the moral dimension are called for. It’s worth emphasizing at the start, then, that the account of mistaken intuitions to be developed

---

71 This is an adapted version of a case introduced by Basu (2019b). The adaption is from Gardiner (2018).
73 These philosophers go on to argue that cases like The Server call for a new account of epistemic rationality that makes room for morality: what is epistemically rational to believe is determined, in part, by moral considerations. This paper, however, will engage with only the negative part of their argument (against evidentialism) rather than the positive part (for moral encroachment).
is not intended to get believers like Spencer off the moral hook. In fact, I will conclude by arguing that – even if the beliefs themselves don’t happen to be prejudiced – such believers may be morally criticizeable in a variety of ways.

1. Prejudiced Belief

It will be helpful to begin by looking at what evidentialists can say about prejudiced beliefs in general, and about the specific beliefs in question in cases like The Server in particular. A few clarifications are in order first. Other than the theoretical assumption that prejudiced beliefs are irrational, I will rely primarily on non-technical, everyday conceptions of prejudice and prejudiced belief. So, nothing more needs to be said about them at a general level upfront.

More does need to be said, however, about the specific beliefs are in question. In cases like The Server, believers have a basic belief that ascribes some property (P) to a targeted social group (G), and a derivative belief ascribing P to an individual member of G (J). For example, Spencer believes that Black diners tip substantially below average (his basic belief) and that Jamal will likely tip below average (his derivative belief). Those who appeal to cases like The Server to challenge evidentialism focus mainly on the derivative beliefs; they will be our main focus, too.

Notice that derivative beliefs are qualified with “likely”. This is important because an unqualified derivative belief automatically goes beyond the evidence. Evidence suggesting that most Gs are P does not firmly suggest that a randomly selected G is P, but it does firmly suggest that a randomly selected G is likely P. So, in order to avoid begging the question of whether the beliefs in question are based on evidence that support them, I focus on qualified derivative beliefs, even though many derivative beliefs out in the world are unqualified.74

With that in mind, let’s turn to what evidentialism can say generally about prejudiced belief. Remember that, according to evidentialism, what is rational to believe is determined only by evidential considerations. If prejudiced beliefs are irrational, evidentialists say that there must be some evidential reason why. On evidentialism, then, it’s essential to prejudiced beliefs that they are not properly based on evidence that supports

---

74 See section 4 of Gardiner (2018) for a helpful discussion of this point.
them. Given this, a natural way for evidentialists to approach prejudiced beliefs is to posit something in their causal history that explains why prejudiced beliefs are irrational. Prejudiced beliefs, evidentialists may contend, are directly caused, at least in part, by some sort of prejudice. And since beliefs directly caused by prejudice can’t be based properly on evidence that supports them, the thought goes, prejudiced beliefs are necessarily epistemically irrational.

This approach allows evidentialists to render the right verdict about beliefs that are caused by internal prejudice. Internal prejudice amounts to something like ill will or a lack of good will towards G and, derivatively, its members, which involves various desires and cares that distort belief formation and maintenance. The sexist’s desire to keep women in their place can cause him to feel threatened by or protective of women which, in turn, can cause sexist beliefs – for example, the belief that women are unfit for high political office despite evidence that clearly suggests otherwise.

This approach also allows evidentialists to render the right verdict about prejudiced beliefs caused by external prejudice. External prejudice amounts to something like culturally prominent stereotypes and evaluative biases about G and, derivatively, its members. These stereotypes and biases can subconsciously distort beliefs when encoded by a believer. Encoded stereotypes about women, for example, may cause an employer to believe that Joe’s application is more impressive than Joanne’s application even though they are equally impressive.

Beliefs directly caused even just in part on internal or external prejudice are not properly based on the evidence, and so evidentialists say they are irrational. Paradigmatic instances of these beliefs fly in the face of widely available evidence, which makes it clear to those of us evaluating them that they likely are not properly

---

75 Arpaly (2003) and Arpaly and Schroeder (2014) develop a volitional-based account of prejudiced belief. They argue that all prejudiced beliefs are based on ill will or a lack of good will. Such accounts echo J.L.A Garcia’s volitional account of racism (which we might extend to other forms of prejudice such as sexism and homophobia). See Garcia (1996, 1997, 1999).

76 Encoded stereotypes and biases might be thought to constitute implicit biases. For a helpful discussion of the epistemic impact of implicit bias, see Gendler (2011) and Egan (2011). For a helpful discussion about the psychology of implicit bias, see Devine (2009).

77 The above characterization of internal and external prejudice is admittedly rough. In real life, the line between them is often blurry, and they interact in complicated ways: internal prejudice can cause and undergird external prejudice and vice versa. Growing up in a community marked by external prejudice against Black people, for example, can foster internal racism against Black people. And internal racism against Black people can lead to continued dissemination of harmful stereotypes and evaluative biases.
based on evidence that supports them. They are also uncharitable, which makes it clear that they are likely the result of internal or external prejudice instead of some other, morally benign distorting influence.

So, on an approach like this, evidentialists can correctly categorize as irrational many important prejudiced beliefs. But what, on this approach, can evidentialists say about beliefs like Spencer’s that seem to be prejudiced but are also backed by the believer’s evidence? Evidentialists must admit that Spencer’s evidence supports his belief about Jamal. After all, statistical evidence that supports the belief that most members of a given group have some property derivatively supports the belief that a randomly selected member of that group likely has that property. But, since it’s possible for a belief to be supported by a believer’s evidence while not being properly based on it, evidentialism can accommodate at least some prejudiced beliefs that are supported by the evidence. These are evidentially-supported beliefs that are caused in part by internal or external prejudice.

Evidentialists can say, then, that there is a possibility that beliefs like Spencer’s are prejudiced even though they are supported by the evidence. But, generally, if it’s known that a belief is supported by the evidence, it’s reasonable to assume that it is properly based on that evidence unless there is good reason to think otherwise. So, unless there is good reason to think that Spencer’s belief is not properly based on his statistical evidence, evidentialism seems to be stuck without much else to say. If the beliefs in question are properly based on the evidence – and perhaps we should think they are – they are rational and, thus, not prejudiced, despite any intuitions to the contrary. But, if they are, instead, prejudiced, then they are not properly based on the evidence, again despite any intuitions to the contrary.

2. An Evidentialist Account of Mistaken Intuitions

This brings us back to the challenge to evidentialism posed by cases like The Server. The challenge starts with the intuitions that the beliefs in question are prejudiced (and, thus, irrational) and also are properly based on the evidence (and, thus, rational). To respond to this challenge, evidentialists must account for these intuitions – that is, they must explain why it seems to so many of us who are evaluating cases like The Server that the beliefs in question are prejudiced and properly based on evidence that supports them.

---

78 In other words, it’s possible for a belief to be propositionally rational without being doxastically rational.
The aim of this section is to develop a plausible evidentialist account of why these intuitions are mistaken. Appealing to the complex socio-epistemic landscape in cases like The Server, I will argue that it is unclear whether the beliefs in question have a certain implicit conjunct, and it is ultimately this unclarity that gives rise to the two conflicting intuitions.

2.1 Two Possible Beliefs

The evidentialist account of mistaken intuitions starts with the claim that beliefs like Spencer’s might have an implicit conjunct. The beliefs in question have the form \( J \) is likely \( P \). The conjunct that’s potentially in play is negatively evaluative and has the form *and this – likely being \( P \) – reflects poorly on \( J \).* So, when looking at beliefs like Spencer’s, there are two possible beliefs it could be: a nonevaluative belief without the implicit conjunct, and a negative belief with the implicit conjunct.79

The beliefs in question have the conjunct if the believers take \( P \) as negative. By “negative properties”, I mean to refer to those properties that reflect poorly on the object to which they are ascribed: a negative property constitutes, signals, or flows from some vice or lack of virtue80 and, thus, the object deserves a low evaluation in light of it. Negative properties, as I take them, are a subset of bad properties. *Being abusive* and *being abused* are both bad in some sense, but only *being abusive* is a negative property. That’s because being abusive reflects poorly on abusers, whereas being abused does not reflect poorly on survivors of abuse.

So, if \( S \) believes that \( J \) is likely \( P \), and if \( S \) takes \( P \) as negative, then \( S \) also (implicitly) believes that likely being \( P \) reflects poorly on \( J \). Applying this to Spencer’s belief: if Spencer takes *likely to tip below average* as negative, he not only believes that Jamal is likely to tip substantially below average, but also (implicitly) that this reflects poorly on Jamal.

---

79 It’s worth noting here that the implicit conjunct may, instead, be thought of as forming a distinct belief without any substantial changes to the account that follows. In this case, the possibilities in play would be Spencer having only one of the two relevant beliefs (the belief that Jamal is likely to tip less than average) and Spencer having both of the relevant beliefs (the belief that Jamal is likely to tip less than average and the belief that *likely to tip less than average* reflects poorly on Jamal).

80 The relevant virtue or vice may be nonmoral.
2.2 The Salience of Both Beliefs

As cases like The Server are usually presented, we do not have enough information to determine whether or not the beliefs in question have the implicit conjunct. Instead, some features of these cases suggest that they do while other features suggest that do not. This renders both possible beliefs – the one with the implicit conjunct, and the one without – salient.

The most obvious feature suggesting that they do not have the conjunct is that the cases explicitly ascribe only the singular belief to the believers. Considering them at face value, then, this suggests that the beliefs in question lack the conjunct. Another such feature is the statistical evidence that is said to support the beliefs in question. Statistical evidence is merely descriptive evidence: it regards only the proportion of Gs who are P. This feature suggests that the beliefs in question are nonevaluative: after all, it is natural to think that the content of Spencer’s belief about Jamal, for example, reflects the nonevaluative nature of the evidence that he has to support it.

But there are other features of these cases that suggest that the beliefs do have the implicit conjunct. Any feature suggesting that the believers take P as negative also suggests that the implicit conjunct is in play. The properties in the cases like The Server – likely to tip less than average or likely belonging to a gang, for example – are often taken to reflect poorly on those to whom they are ascribed. Given this, it is natural for us to think that the believers take them as negative, too.

In addition, some features of these cases suggest, more specifically, that the believers take P to reflect poorly on J qua G. P is taken to reflect poorly on J qua G if P is taken to constitute, flow from, or signal a vice or lack of virtue characteristic of G that J, as a G, likely instantiates. To get a good grasp on the features that suggest this, it will be helpful to have in mind an overview of the socio-epistemic landscape at hand.

The beliefs in question in cases like The Server ascribe P to J, an individual member of G; P is, in fact, common among members of G; and the believers have statistical evidence that supports P’s commonness among G. This all raises the question, why P is common G? In many cases, how we interpret the believers to

---

81 Characteristic features of G are stable and permanent, as opposed to superficial or incidental. They serve to identify G, help distinguish G from other social groups, and are instantiated by typical and, thus, most members of G.
be answering this question informs how we view the beliefs in question. The correct answer to why P is common among G ultimately involves some social injustice suffered by Gs. For example, sociologists typically explain racial disparity in tipping in one of two ways, both of which trace back to anti-Black racism.\textsuperscript{82} The first explanation is that Black people tend to be unfamiliar with percentage-based tipping norms: due to low income levels that result from being subject to a long history of oppression, Black people do not dine at full-service restaurants as much. The second explanation is that servers systematically discriminate against Black diners.

That said, the relevant features of these cases give prominence to a different, mistaken answer to why P is common among G: that P is causally grounded in a vice or lack of virtue characteristic of G. The reason why this answer is salient in these cases has to do with the fact that the social groups targeted are marginalized. In particular, these groups have historically been characterized in negative ways that have been viewed as closely linked to the properties common among them. There is a storied history of thought that maligns Black people, for example, as inherently less virtuous.\textsuperscript{83} Although today’s stereotypes about Black people may be less flagrant, they echo this thought.\textsuperscript{84}

This feature invites us to think that the believers accept the mistaken answer to why P is common G and, by extension, that P reflects poorly on J \textit{qua} G. Given the salience of the negative historical characterization of Black people, for example, we are invited to think that Spencer takes the tipping patterns of Black diners as grounded in something characteristically negative about Black people – specifically, as flowing from a moral defect (lack of generosity, perhaps) characteristic of Black people that Jamal, in turn, likely instantiates.

Notably, something similar can be said about the business consultant who believes that the woman walking out of the corner office is likely an assistant, and the police officer who believes the Hispanic male he sees on the street corner likely belongs to a gang. Given the all-too-familiar characterizations of women as hyper-emotional and irrational, and Hispanic men as violent, we are invited to think that these believers take such negative group characteristics as explaining why women are more likely to be assistants rather than executives.

\textsuperscript{82} See Brewster and Mallison (2009) for a helpful overview and critical discussion of the relevant sociological literature.
\textsuperscript{83} Nussbaum (2015)
\textsuperscript{84} Gendler (2011) cites a 1995 study where participants were asked to name traits that were stereotypically associated with Black Americans. The most typical traits named were negative. See p. 43 of Gendler (2011) for discussion. The original study is from Devine and Elliot (1995).
and why Hispanic men from a certain part of town are more likely than not belong to a gang – though, once again, the correct explanation ultimately traces back to historic and continuing social injustices suffered by these groups.\(^85\) So the fact that the social groups targeted in cases like The Server are marginalized invites us to think not only that the believer takes P as negative, but also as reflecting poorly on J \textit{qua} G.

Another feature that invites us to think the same thing has to do with the context of evaluation, rather than the content of the cases themselves – namely, these cases are typically evaluated within discussions about prejudice. Two things about prejudice are of note here. First, prejudice tends to be negatively-valenced. Second, prejudice is always filtered through a group-level generalization: you can’t be prejudiced against someone \textit{qua} individual, but you can be prejudiced against someone \textit{qua} Black person or woman or Muslim.\(^86\) Relatedly, a hallmark of everyday conceptions of racism and other forms of prejudice seems to be the thought that the targeted group has inherent negative characteristics – vices and lack of virtues – that make that group inferior to other groups in the same social category.\(^87\) Because of this, evaluating the beliefs in question in a context where prejudice in the foreground primes us to think that the believer not only takes P as negative, but also as grounded in a contextually-salient, group-level characteristics.

Given these last two features, it will now be assumed that the implicit conjunct that’s potentially in play has the form \textit{and this – likely being P} – reflects poorly on J \textit{qua} G.

\section*{2.3 The Rise of Conflicting Intuitions}

The salience of the possibility that the beliefs in question lack the implicit conjunct gives rise to the intuition that they are based on the evidence. For example, if Spencer’s belief about Jamal lacks the implicit conjunct, then his statistical evidence that most Black diners tip substantially below average supports his belief. And,

\(^{85}\) See Okin (1989) for an insightful discussion of sexism’s continued impact on women in the workplace. See Knox, Etter, and Smith (2018) for various discussions about the racial-oppression thesis, which says that social injustices suffered by racial minorities are a significant causal determinants of gang membership and activity.

\(^{86}\) This second thing about prejudice is important because we need to be able to distinguish prejudiced attitudes against particular people from other negative attitudes against particular people that arise from personal interactions with them. There is a difference between having ill will towards Joan \textit{qua} woman, for example, and having ill will towards Joan \textit{qua} individual-who-constantly-complains: the first amounts to prejudice, the second does not. See Begby for discussion.

\(^{87}\) This is indicated by standard dictionary definitions of the different forms of prejudice. For example, one of the Oxford English Dictionary’s definitions of racism is “the belief that different races possess distinct characteristics, abilities, or qualities, especially so as to distinguish them as inferior or superior to one another.”
generally, if it is known that a belief is supported by the evidence, it is reasonable to assume that it is based on that evidence, unless there is good reason to think otherwise. Because of this, when we view them as singular and nonevaluative, beliefs like Spencer's strike us being based on the evidence.

The salience of the possibility that the beliefs in question have the implicit conjunct gives rise to the intuition that they are prejudiced. For example, if Spencer's belief about Jamal has the implicit conjunct, it is not supported by the evidence and, thus, cannot be properly based on it. In order for the whole belief to be supported by the evidence, the implicit conjunct must be evidentially supported, but it is not. To see this, it'll be helpful to identify what sort of evidence would be needed in order for the conjunct to be supported. If it is true that \( P \) reflects poorly on \( J \) qua \( G \), then \( P \) is linked to some lack of virtue of vice characteristic of \( G \), and evidence that favors the truth of such a conjunct would suggest so. But Spencer's statistical evidence does not suggest that \( \text{likely to tip less than average} \) is linked to anything negative about Jamal, much less to anything characteristically negative about Black people that Jamal likely instantiates.

At the very least, then, we have reason to think that, if they have the implicit conjunct, the beliefs in question go beyond the evidence in an uncharitable direction. But we might also have reason to think that they go against it.\(^88\) If we have been envisioning Spencer in a multicultural society where he regularly interacts with Black people – they are his neighbors and friends and colleagues – then we have reason to think that he has plenty of evidence to counter the thought that Black people are, for example, characteristically less generous.

Whether these beliefs go beyond or against the evidence, they are not supported by the evidence and, thus cannot be properly based on it. And since they do so in an uncharitable direction, given the evidentialist approach discussed above, we have good reason to think that they have been influenced by prejudice. Because of all of this, when we view them as having the implicit conjunct, beliefs like Spencer's strike us as being prejudiced.

---

\(^88\) S's belief goes beyond the evidence if S's evidence favors, on balance, the truth of its content but not to the degree required for rationality, or if S's evidence favors, on balance, the truth of some of its content and neither favors nor disfavors the rest. S's belief goes against the evidence if S's evidence favors, on balance, the negation of part or all of its content.
2.4 The Account

With all this in mind, an account of mistaken intuitions can now be offered. There is a fact of the matter about the content of the beliefs in question: they are either nonevaluative beliefs that lack the implicit conjunct or negative beliefs that have it. But, given the limited information presented in The Server and cases like it, those of us evaluating the beliefs in question are not in an epistemic position to identify what these facts are. Rather, the possibility of each belief is salient.

As we evaluate the beliefs in question, we do not consciously disambiguate these possibilities – for example, we do not think about Spencer’s belief about Jamal as possibly a singular, nonevaluative belief or possibly a conjunctive, negative belief. Rather, we just think about it simply as Spencer’s belief, its content solidified. We do, however, register the various features that favor the different possibilities, leading us either to subconsciously alternate between viewing Spencer’s belief as singular and nonevaluative and then viewing it as conjunctive and negative, or to subconsciously view it as both simultaneously. Viewing it the first way gives rise to the intuition that it is based on the evidence and viewing it the second way gives rise to the intuition that it is prejudiced. Since we explicitly think about it as being one belief, we incorrectly take both intuitions as being about one belief – Spencer’s belief about Jamal – not as intuitions about different possible beliefs.

2.5 The Upshot

Suppose this account plausible. How, exactly, does it help out evidentialism? Evidentialism cannot accommodate beliefs that are both prejudiced (and thus irrational) and properly based on evidence that supports them (and thus rational). But beliefs like Spencer’s intuitively seem to be what evidentialism says is impossible. If both of these intuitions are veridical, then evidentialism is incorrect, and a new account of the rationality of belief is needed that can accommodate these beliefs. But if these intuitions are mistaken – if it’s not the case that the beliefs in question are prejudiced and also properly based on the believers’ evidence – then evidentialism can accommodate them. And so, in the face of cases like The Server, an account of mistaken intuitions makes it reasonable for evidentialists to hang on to their view about epistemic rationality.
3. Case Variations: Testing the Account

If the account of mistaken intuitions offered above is correct, our intuitions about the beliefs in question will differ in variations of the cases at hand in which there are not mixed messages about their content – in other words, variations in which it is clear whether or not it’s implicitly believed that P reflects poorly on J \textit{qua} G. Consider the following two variations on The Server.

**The Informed Server** Spencer knows that, statistically, Black diners tip substantially below average and derivatively comes to believe that Jamal will likely tip below average. Spencer has recently read a lot about the historic and continued oppression of Black Americans. So, in addition to the relevant statistical information, he knows that they have been disadvantaged by structural racism for centuries and that this has led to systematic income inequality between Black American and White Americans. From his research, he knows that it is this income inequality, and not any lack of virtue, that ultimately explains the tipping patterns of Black Americans.

**The Ignorant Server** Spencer knows that, statistically, Black diners tip substantially below average and derivatively comes to believe that Jamal will likely tip below average. Despite daily interactions with his Black neighbors and coworkers that provide him with plenty of evidence to the contrary, Spencer believes that Black people are inherently less morally virtuous than White people. He believes that it is a lack of generosity that explains the tipping patterns of Black diners.

In both variations, Spencer believes that Jamal will likely tip below average. The beliefs that surround the belief in question, however, differ drastically from one variation to the next. The surrounding beliefs make it clear in each variation whether or not Spencer takes \textit{likely to tip below average} as reflecting poorly on Jamal \textit{qua} Black person: the features of the original case that suggest one way are underscored, and any remaining features that suggest the other way are overshadowed.

In The Informed Server, the surrounding beliefs make it clear that Spencer does not take \textit{likely to tip below average} as flowing from something negative about Jamal as a Black person, and dispel any hint that he sees Black people in the negative light that cast by any contextually salient history of racism. In The Ignorant Server, the surrounding beliefs make it clear that Spencer takes \textit{likely to tip below average} as flowing from a moral defect characteristic of Black people. It is clear, then, that the belief in question has the implicit conjunct in The Ignorant Server, but not in The Informed Server. By extension, it is clear that the other salient possibility – that it lacks the implicit conjunct – obtains in The Informed Server and not in The Ignorant Server.

These variations can serve as a confirmation test for the account of mistaken intuitions developed above. The account is confirmed if we have the intuitive responses that the account predicts for each variation. In the
Informed Server, the only salient possibility is the singular, nonevaluative belief. So, regarding Informed Spencer’s belief, the account predicts that we retain the intuition that it is properly based on his evidence and lose the intuition that it is prejudiced. In the Ignorant Server, the only salient possibility is the conjunctive, negative belief. So, regarding Ignorant Spencer’s belief, the account predicts that we retain the intuition that it is prejudiced but lose the intuition that it is based on the evidence. And this is exactly how things seem. Informed Spencer’s belief does not seem racist but does seem to be epistemically in-the-clear by evidentialist lights. By contrast, Ignorant Spencer’s belief does seem racist but also seems epistemically problematic, stemming from the type of unsupported ideological commitments that tend to go hand-in-hand with internal racism.

These intuitive responses are enlightening. Of course, they confirm the account of mistaken intuitions. But, in so doing, they indicate something else important: just believing that Jamal will likely tip less than average is not, itself, automatically morally pernicious. If it were, then Informed Spencer’s belief about Jamal would also seem racist. This, in turn, indicates that there’s more going on than first meets the eye in the original version of The Server that is triggering our intuition about racism, which is exactly what that account of mistaken intuitions seeks to identify – namely, tacit features suggesting that Spencer’s belief has the implicit conjunct. It is by appealing to the different ways that these features as well as the relevant competing features are developed in each variation that evidentialists can predict – and explain – our differing intuitive responses. Because the features of these variations point univocally towards only one of the possible beliefs, we view the belief in question only through the lens of that possibility, which gives us the intuition that arises from viewing it in that way. Considerations about these variations, then, strengthen the account of mistaken intuitions.

4. Some Concluding Considerations: The Moral Dimension

We began with the intuitions that Spencer’s belief about Jamal in the original version of The Server is prejudiced and also that it is properly based on evidence that supports it. Evidentialism cannot accommodate prejudiced beliefs that are properly based on the evidence because, on evidentialism, such beliefs would be both rational and irrational. But evidentialists can offer an account that explains away these intuitions. If plausible, this
account disarms the challenge to evidentialism posed by cases like The Server. And we do have some good reason to think it’s plausible since it predicts and explains intuitive verdicts about relevant variations on The Server.

One of the virtues of the account of mistaken intuitions is that it is sensitive to the complexity of the socioepistemic dimension of cases like The Server. In concluding, I want to emphasize the complexity of their moral dimension. This emphasis is important because it shows that – in addition to accounting for the mistaken intuitions – evidentialism can accommodate a robust and nuanced moral evaluation of these cases.

To start, it should be noted that cases like The Server have many morally significant aspects. Because of this, even if it were determined that the beliefs in question are properly based on the evidence and so are not prejudiced, believers like Spencer are not automatically in the moral clear. There are a number of moral questions whose answers contribute to a full moral evaluation of cases like The Server, but that are not settled by determining whether the beliefs in question are prejudiced. Let’s explore two such issues: actions that may stem from the beliefs in question and attention directed towards their content.

When considering The Server, it’s natural to wonder why Spencer occurringly comes to believe that Jamal is likely to tip less than average. One obvious reason might be that it’s relevant to a decision about how to act. But, in The Server, it’s stipulated that Jamal is sitting outside of Spencer’s section. This stipulation was included because it makes it even harder for evidentialists to account for the mistaken intuitions – evidentialists can’t try to explain away the intuition that the beliefs in question are prejudiced (or, more generally, morally problematic) by pointing to morally condemnable actions that the belief would likely cause. After all, if Spencer won’t be interacting with Jamal, then his belief that Jamal will likely tip less than average won’t be informing any relevant actions.

The main point here is not that, in a case without this stipulation, potential actions downstream from Spencer’s belief can bolster the evidentialist account of mistaken intuitions. Rather, the main point is that these potential actions are morally significant in themselves: when the beliefs in question seem to recommend a course of action that risks harm to or disadvantages J, the moral status of the beliefs themselves – whether or not they are prejudiced – does not tell us whether it is morally permissible for the believer to take that course.
of action. So, even if Spencer’s belief about Jamal is not prejudiced, and supposing Jamal is sitting in Spencer’s section, it may still be morally wrong for Spencer to give Jamal poorer service than he would give to a diner whom he does not believe is likely to tip less than average.

For another example, consider an HR manager who believes that Jenny – a young married female applicant whom he is interviewing – is likely to become pregnant soon after being hired in light of statistical evidence that most young married female applicants he’s hired in the past have become pregnant in their first five years of employment.\footnote{This example is adapted from a case discussed in Schroeder (2018).} Even if it the manager’s belief lacks the implicit conjunct, and is properly based on his evidence and so is not prejudiced, it may still very well be morally wrong for him to count statistically likely to become pregnant soon after being hired as a strike against Jenny when deciding whom to hire.\footnote{For empirical studies on pregnancy discrimination in hiring, see Cunningham and Macan (2007) and Becker, et al., (2019).} This shows that evidentialists have room to morally criticize some believers like Spencer even on the assumption that the beliefs in question are epistemically and morally in the clear.

Now let’s return to cases where the belief in question does not seem likely to lead to a course of action that risks harm to J – cases like The Server in which Jamal is sitting outside of Spencer’s section. For another example, suppose instead it’s a low-level employee at the company – somebody with no say in the hiring process – who comes to believe in light of the relevant statistical information that Jenny, if hired, is likely to become pregnant soon after. Although morally problematic potential actions are not relevant in such cases, another moral issue is: the believers’ attention. Supposing that Jamal’s likely tipping proclivities are totally irrelevant to deciding how to act, why is Spencer thinking about this? Why does he seem to care about how much Jamal is likely to tip? It may seem odd that Spencer is seemingly so interested in the tipping patterns of Black diners that this is what comes to mind when he sees Jamal. Sure, there might be an innocuous explanation – for example, maybe Spencer just studied differing social customs among racial groups in his sociology class. But other, less innocuous explanations involving racist ideologies or attitudes seem more likely, or at least salient. And something similar can be said about the low-level employee.
Again, the main point here is not that, in cases where potential morally condemnable actions downstream from the belief in question are not in play, evidentialists can bolster their account of mistaken intuitions by pointing to morally fishy attention patterns involved in occurrently having the beliefs in question. Rather, the main point is that these attention patterns are morally significant in themselves. Once again, this leaves evidentialists with room to morally criticize some believers in cases like The Server even on the assumption that the beliefs in question are not prejudiced. For example, even if Spencer’s belief about Jamal itself isn’t prejudiced, his attention patterns – manifested in him occurrently having that belief when its content isn’t practically relevant – may nevertheless stem from racist attitudes. In this case, Spencer’s attentions patterns, but not his belief, would be racist. In addition, there may be moral norms that govern attention which, by occurrently having the beliefs in question, the believers violate. Perhaps, all else being equal, it’s morally wrong to dwell on things about a person that are morally and practically irrelevant but incline us to think worse of that person, distract us from things that are relevant, or feed into false or harmful stereotypes about the groups to which that person belong.

In sum: cases like The Server are complicated, both morally and epistemically. Initial intuitions about the beliefs in question in these cases pose a challenge to evidentialism. But evidentialists can successfully respond to this challenge by developing an account of mistaken intuitions: conflicting features of such cases as originally presented muddy our intuitions about the beliefs in question. This account allows evidentialists to explain and predict intuitive responses to variations of these cases in which its clear which of these features hold; this, in turn, lends farther plausibility to the evidentialist diagnosis of what’s going on in the original cases. What’s more, evidentialism can support a sophisticated evaluation of the moral dimension of these cases over and above its account of the epistemic dimension. So, far from pressuring evidentialists to give up their view of epistemic rationality, cases like The Server fit comfortably within an evidentialist framework.

---

91 For one interesting discussion of the moral significance of attention, see Bommarito (2013) who argues that attention patterns are central to the virtue of modesty and the vice of immodesty.
Works Cited


Chapter 4
Prejudiced Belief and the Moral Importance of Understanding
A Challenge to Revisionist Views of Epistemic Rationality

Recently, there has been a lot of attention in epistemology given to cases involving beliefs that seem to be prejudiced but also properly based on evidence that supports them. These cases all share a similar structure: a believer (S) has strong statistical evidence that most members of a certain marginalized social group (G) have some property (P) and then comes to believe that an individual (J) who is a member of G is likely P. Here is one such case.

**The Server** Spencer works as a server at a restaurant. He sensed that White diners tipped more than Black diners. Doing a bit of research online, Spencer found a well-documented social trend that Black diners tip substantially below average. Spencer weighs the evidence before reaching his belief that Black diners tip substantially below average. A Black diner, Jamal, enters Spencer’s restaurant and dines in a booth outside of Spencer’s section. Spencer believes that Jamal will likely tip below average.92

There are many other cases like The Server. For another example, consider a school teacher who, upon meeting two new students – Jenna and Joel – comes to believe that Jenna likely scored lower than Joel on last year’s statewide standardized math exams given evidence that girls tend to score lower than boys on such exams.

Cases like these have garnered so much interest in epistemology, in part, because they raise important questions about epistemic rationality. According to a popular evidentialist view – call it *evidentialism* – epistemic rationality is determined alone by evidential considerations: S’s belief is epistemically rational iff S’s evidence supports its propositional content and it is based properly on that evidence.93 Assuming that prejudiced beliefs are irrational,94 evidentialism entails that there aren’t any prejudiced beliefs that are properly based on the evidence. But that’s exactly what the beliefs in question in cases like The Server seem to be. There seems to be

---


93 Three terminological notes are worth making here. First, “evidence” is meant broadly to include all truth-related considerations, whether or not they are strictly evidential. Second, this account of epistemic rationality concerns doxastic – rather than propositional – rationality. Third, “rationality” is meant to refer to epistemic rationality from here on.

94 This is a common assumption about prejudiced beliefs that’s shared by many evidentialists and revisionists alike. This assumption will be taken as a given in this paper.
something racist about Spencer’s belief about Jamal, for example, but it also seems to be backed by Spencer’s statistical evidence.

Some philosophers – call them revisionists – appeal to these cases not only to argue against evidentialism, but also to motivate a new view of epistemic rationality. Specifically, they argue for moral encroachment, which is that view that what is rational to believe is determined, in part, by moral considerations. Revisionists agree with evidentialists that there are no prejudiced beliefs that are rational, but that’s because of their revisionary account of rationality, not because there are no prejudiced beliefs that are properly based on supporting evidence. In fact, there are prejudiced beliefs that are properly based on supporting evidence: beliefs like Spencer’s.

My goal here is to address the positive part of the revisionist project (for moral encroachment), rather than the negative part (against evidentialism): upon closer inspection cases like The Server do not motivate moral encroachment as clearly or cleanly as revisionists suggest. Towards this goal, I consider variations on these cases in which the beliefs in question are embedded in an understanding of morally relevant features of the believers’ socio-epistemic landscape. The revisionist approach to cases like The Server, I argue, is committed to an implausible evaluation of these variations, and this suggests that the revisionist approach is too strong. What’s more, I argue, an evidentialist approach has room to accommodate a more plausible evaluation of these variations. Along the way, I’ll explore the moral importance of understanding, as well as the complexity of the socio-epistemic landscape that believers like Spencer find themselves in today. In light of all this, I conclude by gesturing at a way that evidentialists can address the negative part of the revisionist project.

1. The Two Approaches

In order to get a handle on why cases like The Server are thought to motivate revisionism, it’s helpful to look first at why evidentialism seems to fall short. To start, then, let’s look at both the evidentialist and revisionist approaches to the particular beliefs in question in these cases, as well as to prejudiced beliefs more generally.

Before moving forward, I need to make a few clarifications. First, I will be relying on common, everyday conceptions of prejudice and prejudiced belief, with one exception for the theoretical assumption that
prejudiced beliefs are epistemically irrational. Prejudice is meant as a general category which includes a variety of types of prejudice such as racism, sexism, anti-Semitism, and homophobia. Accordingly, prejudiced beliefs include, among others, racist beliefs and sexist beliefs.

Second, it’s worth getting clear on which particular beliefs are in question. In cases like The Server, the believers have a basic belief that ascribes P to G, and a derivative belief ascribing P to J, an individual member of G. For example, Spencer believes that Black people tip substantially lower than average (his basic belief) and that Jamal will likely tip less than average (his derivative belief).\(^95\) Revisionists focus primarily on the derivative beliefs, and so they will be my primary focus, too.

1.1 The Evidentialist Approach

In conjunction with the assumption that prejudiced beliefs are epistemically irrational, evidentialism entails that there are no prejudiced beliefs that are properly based on supporting evidence. Because of this, a natural way for evidentialists to approach prejudiced beliefs is to presuppose something in their causal history that ensures that they are not properly based on supporting evidence. Evidentialists might argue, for example, that prejudiced beliefs must be based in part on prejudice, which, in turn, prevents them from being properly based on supporting evidence.

An approach like this allows evidentialists to account for a number of prejudiced beliefs, including those based on internal prejudice or external prejudice. By internal prejudice, I mean negative affective and motivational attitudes a person harbors towards G and, derivatively, its members.\(^96\) By external prejudice, I mean something like culturally prominent negative stereotypes and evaluative biases about G and, derivatively, its members. These stereotypes and biases can then be encoded by believers, sometimes unconsciously.

\(^95\) Notice that the derivative beliefs in question are all qualified with “likely”. This is important because an unqualified derivative belief is not supported by statistical evidence. After all, evidence that most Gs are P does not firmly suggest that a randomly selected G is P, but it does firmly suggest that a randomly selected G is likely P. So, in order to clearly leave open the possibility that the beliefs in question are properly based on supporting evidence, I focus on qualified derivative beliefs, even though many derivative beliefs had by real-life believers are unqualified. See section 4 of Gardiner (2018) for a discussion of this point.

\(^96\) See the work of J.L.A Garcia for a related volition-based account of racism.
Internal prejudice and encoded external prejudice can both illicitly influence beliefs, either by leading to beliefs that are not supported by the evidence or by preventing beliefs that are supported by the evidence from being properly based on that evidence. For example, the racist’s desire to keep Black people in their place can cause her to form wish-fulfilling beliefs – such as the belief that Black men are dangerous – which, in turn, justify her efforts to do just that. For another example, encoded stereotypes about women can cause a board member to believe that the woman candidate doesn’t have what it takes to be CEO, despite her excellent track record in executive roles.

On the approach on offer, prejudiced beliefs tend to have two characteristics that enable us to identify them as prejudiced. First, prejudiced beliefs tend to go beyond or against the evidence we know or expect the relevant believer to have. Second, prejudiced beliefs skew negative, which indicates that they have likely been illicitly influenced by prejudice rather than some other, less nefarious distorting influence.

This evidentialist approach can correctly categorize many prejudiced beliefs. But what can evidentialists say about beliefs like Spencer’s that seem to be prejudiced but also backed by relevant statistical evidence? At first, it seems, not much. Generally, if it’s known that a belief is supported by a believer’s evidence, it’s reasonable to assume that it’s also properly based on the evidence, unless there is good reason to think otherwise. So, unless there is good reason to think that Spencer’s belief is not based on his statistical evidence, evidentialists are still faced with explaining the beliefs in question in cases like The Server. If they are properly based on the statistical evidence – and, at least at first glance, we don’t have good reason to think they’re not – these beliefs are neither irrational nor prejudiced, despite intuitions to the contrary.

This is what leads revisionists to call out for a new account of epistemic rationality that can accommodate prejudiced beliefs that are based properly based on supporting evidence.
1.2 The Revisionist Approach

At the heart of the revisionist approach to cases like The Server is a view that has received a lot of attention lately: moral encroachment. According to moral encroachment, morality gets a direct say in what is rational to believe. Specifically, moral considerations determine how much, or what type of, evidence is needed for a belief to be rational. When the moral stakes for having a certain belief are high, morality sets a high evidential threshold for that belief. When the moral stakes are low or absent, morality sets a lower evidential threshold. This allows revisionism to say that it is rational to believe, for example, that a randomly selected canary is likely to be yellow given evidence that most canaries are yellow, but it is not rational to believe that a randomly selected Black diner likely tips below average given evidence that most Black diners tip below average, even if both beliefs are epistemically flawless according to evidentialism.

Like evidentialists, revisionists easily render the right verdict about paradigmatic prejudiced beliefs: the evidence is decidedly not in their favor and so these beliefs are clearly irrational—they wouldn’t even pass a low evidential threshold.

But revisionists also seem to have something compelling to say when it comes to beliefs like Spencer’s. To see this, it’s helpful to look at how revisionists figure out the moral stakes of having certain beliefs. Most revisionists agree that risk of harm plays an important role: when a belief carries a significant risk of harm, the moral stakes are high, and so more (or better) evidence is required than otherwise would be for that belief to be rational. They then argue that derivative beliefs like the ones in question risk a lot of harm—that is, these beliefs carry a high chance of harm and the chanced harm is significant, especially since the targeted social group is marginalized. Finally, they argue that the believers’ statistical evidence is not (good) enough to push the beliefs over the resulting high evidential threshold.

There are various suggestions about what harms, exactly, are risked by such beliefs. Primarily, these are harms to the individual who is the object of belief, but some are also harms to the targeted social group. Bolinger (2020) suggests that, if the individuals do not have the property that they are predicted to have, such beliefs

---

97 For defenses of moral encroachment, see Basu (2018c), Basu and Schroeder (2019), Fritz (2017), Moss (2018), and Pace (2011). Bolinger (2019) argues for moral encroachment on rational acceptance and remains neutral on whether acceptance and belief are the same attitudes. See Gardiner (2018) for a critical discussion.
harm by offending the individual’s status. If Jamal prides himself in leaving large tips, for example, Spencer’s belief offends his status as a generous tipper. Basu (2019a) also suggests that such beliefs harm by violating the individual’s general interest in being treated as a person rather than an object. Being treated as a person, she argues, has a doxastic component: for example, the school teacher’s belief about Jenna student should be sensitive to how Jenna is personally, and not how she is predicted to be in light of her gender. Moss (2018) and Schroeder (2018) both suggest that such beliefs harm by leading to actions that harm the individual. For example, the school teacher’s belief about Jenna student might lead her, perhaps subconsciously, to overlook or fail to foster Jenna’s mathematical talent and interests. And Basu (2018a and 2018c), Bolinger (2020), and Moss (2018) all suggest that such beliefs contribute to collective harms from which the individual and targeted social group suffer. For example, many servers believing that Black diners tip substantially below average leads to systematically poor service to Black diners which, in turns, discourages Black patronage, exacerbating the ills of social segregation; such beliefs being prevalent also makes it harder for restaurants to retain servers in areas with a large percentage of Black patrons, which makes owners averse to opening restaurants in Black communities.

Because the beliefs in question risk a lot of harm, they face a high evidential threshold: they require a lot of (good) evidence to count as rational. The sort of statistical evidence that is sufficient to rationalize low-stakes derivative beliefs about canaries, revisionists argue, is not sufficient to rationalize high-stakes derivative beliefs like Spencer’s. Beliefs like these must be based on evidence specific to the individual if they are to be rational: Spencer’s belief that Jamal is likely to tip below average would need to be based on evidence about Jamal himself. Since Spencer does not have such individual-specific evidence, his belief about Jamal does not meet the evidential threshold, which makes it irrational – even if it properly based on Spencer’s statistical evidence.

---

98 Along similar lines, Basu (2018a) suggests that they risk challenging the central self-descriptions through which individuals find their self-worth.
99 For a thorough treatment of the ills of social segregation, see Anderson (2010).
101 Revisionists sometimes insist that, according to their standards, beliefs like Spencer’s are not sufficiently supported by his evidence. See, for example, Basu and Schroeder (2018). But revisionists agree that they are sufficiently supported by his evidence according to evidentialist standards. Revisionists, then, agree that these beliefs sufficiently supported by the evidence according to the standards that concern us.
What’s more, revisionists add, having a belief that doesn’t meet an evidential threshold set high by morality entails that both belief and believer are morally condemnable. Given the high moral stakes, morality demands “extra care and diligence”\textsuperscript{102} when it comes to these beliefs. This amounts to a moral demand to have such beliefs only if there is a lot of good evidence in their favor. By not meeting the relevant evidential threshold, Spencer’s belief flouts this moral demand, and, by dint of having the belief, so does Spencer himself. Spencer wrongs Jamal by virtue of believing as he does.

But just because a belief is irrational and morally condemnable does not entail that it is prejudiced: being irrational and morally condemnable are necessary but not sufficient conditions for prejudiced belief. In addition, prejudice must somehow be involved in a prejudiced belief’s causal history. According to the evidentialist approach, prejudiced beliefs must be directly based in part on internal prejudice or encoded external prejudice. The revisionist approach also counts these beliefs as involving prejudice in the necessary way, but must also count more – after all, revisionists want to count as prejudiced beliefs like Spencer’s even if they are properly based on supporting evidence. So the revisionism approach says that there is an additional, looser way for prejudice to cause beliefs.

Revisionists are not explicit about what this way is, but they do leave hints. Consider a point frequently raised by revisionists: the world itself is racist and sexist – it “has been, and continues to be structured by racist [and sexist and so on] attitudes and institutions…[and so] the evidence might be stacked in favor of racist beliefs.”\textsuperscript{103} The suggestion here seems to be that there are prejudiced parts of the world and so a belief that accurately represents a prejudiced part of the world may, itself, be prejudiced, so long as it meets the other conditions of prejudiced belief. If that’s right, then, on the revisionist approach, a belief’s causal history involves prejudice if prejudice causes something to be the case and, thus, causes there to be evidence that it is the case, and this evidence goes on to be the proper basis of belief. Spencer’s belief about Jamal is caused by prejudice in this way, for example, if it is because of racism that it is true – and, thus, there is evidence – that Black diners tip substantially below average. As will be discussed in section 3, this is exactly how things are in the cases at

\textsuperscript{102} Basu (2019c), p. 20
\textsuperscript{103} Basu (2019b), p. 2498
hand. So the revisionist approach can deem beliefs like Spencer’s as not only as morally condemnable, but also as downright prejudiced.\textsuperscript{104}

### 2. Case Variations: A Challenge for Revisionism

With a firm grasp on both the evidentialist and revisionist approaches to cases like The Server, we are now in a position to challenge the contention that the revisionist approach is better. The challenge I want to develop stems from considerations about variations of these cases in which the beliefs in question are embedded in an understanding of the socio-epistemic landscape. Consider the following variation on The Server.

#### The Informed Server

Spencer knows that, statistically, Black diners tip substantially below average and derivatively comes to believe that Jamal will likely tip below average. Spencer has recently read a lot about the historic and continued oppression of Black Americans. So, in addition to the relevant statistical information, Spencer knows that Black Americans have been disadvantaged by structural racism for centuries and that this has led to systematic income inequality between Black American and White Americans. From his research, he knows that it is this income inequality, and not any lack of virtue, that ultimately explains the tipping patterns of Black Americans.

In sharp contrast with Spencer’s belief about Jamal in the original case, Informed Spencer’s belief about Jamal seems both epistemically and morally in-the-clear. There may even be something morally commendable about Informed Spencer and his doxastic state. But the same revisionist reasoning used to show why original Spencer’s belief gets a failing grade also applies to Informed Spencer’s belief. So, the revisionist approach seems to be committed to the wrong verdicts here: it says that beliefs like Informed Spencer’s are irrational, morally condemnable, and prejudiced, and also that Informed Spencer himself is morally condemnable.

Let’s start by looking more closely at the revisionist verdicts about the beliefs in question in the variation before turning to its verdict about the Informed Spencer himself. Remember that, on the revisionist approach, if a belief about an individual carries a decent chance of non-negligible harm to that individual, morality sets a high evidential threshold for that belief, and statistical evidence is not good enough to push it over. If this belief

\textsuperscript{104} To clarify, a belief being caused by prejudice in this way does not entail that it is prejudiced – to count as prejudiced, such beliefs also have to be irrational and morally condemnable by dint of failing to meet an evidential threshold set high by morality. Consider my belief simply that anti-Black racism still runs rampant in Western societies today. Although this belief involves prejudice in the relevant way, it doesn’t meet the other necessary conditions of prejudiced belief so is not itself prejudiced.
is held nevertheless, it is irrational. It's also morally bad: morality requires that we handle beliefs that risk harm with “extra care and diligence” and beliefs that fail to meet an evidential threshold set high by morality fail to meet this requirement. In addition, these beliefs are downright prejudiced if prejudice is involved in their causal history.

In light of this, where does Informed Spencer’s belief about Jamal stand? Well, it’s plausible that Informed Spencer’s belief risks non-negligible harm to Jamal even though it is embedded in the relevant sort of understanding. And, because there remains a risk of harm, morality sets a relatively high evidential threshold for Informed Spencer’s belief. But Informed Spencer does not have good enough evidence to push it over. So Informed Spencer’s belief is irrational according to the revisionist approach. It is also morally condemnable since it flouts morality’s demand for “extra care and diligence.” What’s more, it involves prejudice in one of the necessary ways since – as will be detailed below – it is ultimately because of racism that Black diners tip less than average.

But, again, these seem like all the wrong verdicts: beliefs like Informed Spencer’s seem to be morally and epistemically in-the-clear.

Let’s now turn to what the revisionist approach entails about Informed Spencer himself. We’ve just seen that, on the revisionist approach, Informed Spencer’s belief fails to satisfy a moral demand for “extra care and diligence”. This entails that, by dint of having the belief in question, Informed Spencer himself fails to satisfy a moral demand. Because those who fail to satisfy a demand of morality are morally condemnable for so failing, Informed Spencer is morally condemnable by virtue of believing that Jamal is likely to tip less than average given his statistical evidence.

Basu (2019c) expounds on the idea that believers like Spencer are, themselves, morally condemnable by offering a revisionist picture of “being woke” that implies that the moral failure of both original Spencer and

105 In the following section, it will be argued that the relevant sort of understanding does significantly mitigate the harm risked to the objects of the beliefs in question. Even so, it does not erase the risk of harm.

106 The term “being woke” and its variants originate within African-American culture, and has grown in prominence throughout the Black Lives Matter movement. Although I am concerned about terminological appropriation, I go ahead use the term here since it is already in use in the literature and since there is no other available term that adequately picks out the intended referent. I put it in quotations to acknowledge that I am borrowing the term.
Informed Spencer is, ultimately, a failure to “be woke.” “Being woke” involves a morally significant cognitive (doxastic) response to the social injustices, both local and structural, in a believer’s immediate socio-epistemic landscape and broader culture. Basu argues that “being woke” amounts, specifically, to belief formation and maintenance that is sensitive to the moral stakes of having certain beliefs in the ways that revisionism spells out. She writes, “we [can] understand moral encroachment as a systematic treatment of the imperative to stay woke.” “Woke” believers do not have beliefs that risk harm to their objects without a lot of (good) evidence in their favor; they do neither form nor maintain derivative beliefs on the basis of statistical evidence, especially when the targeted social group is marginalized. “Being woke”, according to this revisionist picture, is a moral obligation – one that original Spencer and Informed Spencer alike fail to satisfy.

The verdict that Informed Spencer himself is morally condemnable runs against the intuition he is not. This intuition likely has a plausible source – namely, the conviction that believers can’t be morally condemnable just for believing something because of the good reasons they have for thinking it’s true. It also may seem that understanding featured in cases like The Informed Server is the very sort of doxastic response that is characteristic of “wokeness.”

All of this points to a challenge for the revisionist approach: in their attempt to condemn some prejudiced beliefs and believers that the evidentialist approach seems to condone, revisionists mistakenly condemn the innocent.

3. Understanding and Its Moral Importance

The intuitive reaction to The Informed Server highlights an important point about the beliefs in question. As Gardiner (2018) puts it, the “moral character” of such beliefs depends on the understanding in which they are embedded. The challenge for the revisionist approach is that it’s committed to an implausible evaluation of cases like Informed Server in which believers understand important features of their socio-epistemic landscape. Can the evidentialist approach do any better with regard to these variations? I will argue that it can. More

---

107 S’s understanding of x is the network of related beliefs about x that S has. Following Kvanvig (2003, 2009), Pritchard (2007), Hills (2009), and others, it will be assumed that understanding is factive – believers can have misunderstandings, but they cannot have an inaccurate understanding. This assumption is not substantial; it is made for the sake of terminological clarity.
specifically, I will argue that, by appealing to the moral importance of understanding, the evidentialist approach can accommodate a plausible evaluation of cases like The Informed Server. To pave the way for this argument, we need to identify the content of the relevant sort of understanding in these cases at hand. We also need to identify why it is morally important. This, in turn, requires us first to get a handle on the socio-epistemic landscape in play.

3.1 The Socio-Epistemic Landscape

In the cases under consideration, the beliefs in question ascribe P to J, an individual member of G; P is, in fact, common among members of G; and the believers have statistical evidence that supports this. In addition, G is a marginalized group – that is, G has been historically been characterized in negative ways that have been appealed to explain why certain properties are common among members of G. For example, there is a long history of thought that disparages Black people as inherently less virtuous. Today’s stereotypes about Black people are informed by this negative characterization.

These features of the socio-epistemic landscape render salient the question of why P is common among Gs. What’s more, these features put believers at risk of answering this question incorrectly. Specifically, believers are at risk of taking P as causally grounded in something characteristically negative about G – a vice or lack of virtue characteristic of G – and then appealing to this to explain why P is common among G, and, by extension, why J is likely P. For example, Spencer may take tipping less than average to flow from a lack of generosity characteristic of Black people. Similarly, the school teacher may think that scoring lower on math exams flows from irrationality or lack of analytical prowess characteristic of girls and women. After all, these are the answers that are recommended by the relevant culturally prominent narratives of these groups, and the corresponding negative characterizations.

---

108 Nussbaum (2015)
109 See Gendler (2011) for a discussion of a 1995 study where participants were asked to name traits that were stereotypically associated with Black Americans. The most typical traits named were negative. See p. 43 of Gendler (2011) for discussion. The original study is from Devine and Elliot (1995).
110 The relevant vice or virtue may be non-moral.
But, as noted, this is an incorrect answer: in the cases at hand, P is not common among G because P is causally connected to a vice or lack of virtue characteristic of G. The correct answer to why P is common among G traces back to a social injustice suffered by Gs. For example, there are two main sociological explanations of racial disparity in tipping that both trace back to anti-Black racism. The first explanation is that, because of low levels of income that result from being subject to anti-Black racism, Black people do not dine out at full-service restaurants as much and so are not as familiar with percentage-based tipping norms. The second explanation is that Black diners are systemically discriminated against by servers. On both explanations, it is because of racism, and not because a vice of lack of virtue characteristic of Black people, that Black diners tip substantially less than average.

Something similar can be said about why girls and women tend to score lower on math exams than boys and men – and it’s not because girls and women are inherently less analytical or irrational. Rather, empirical studies suggest that stereotypes related to gender and mathematical ability positively affect boys’ performance and negatively affect girls’ performance in competitive environments when it comes to math.

Before moving forward, it’s worth noting that this sort of socio-epistemic landscape often cultivates a vicious feedback loop: when P’s commonness among G is a result of social injustice suffered by G, and there is a salient social narrative about G that primes believers to take P as connected to something characteristically negative about G, believers taking P in this way tends to prop up the relevant prejudice and its constitutive social injustices. For example, racism causes Black people to tip less than average → people believe that Black people tip less than average → some (mis)take this to confirm the view that Black people are less virtuous → this view justifies and, thus, perpetuates the racism that causes Black people to tip less than average.

111 Importantly, I am not claiming here that there are no cases in which P is common among G because P is causally rooted in something characteristically negative about G. Consider a case in which a believer ascribes a property such as likely to oppose measures to make voting easier to some individual member of the KKK. Given the racist attitudes and ideological commitments characteristic of the KKK, the believer has good reason to think that this property is, in fact, linked to some vice or lack of virtue characteristic of the KKK that this individual likely instantiates. Cases like this show that it may sometimes be reasonable to think that it’s because of something negative about G that J is likely P. But such cases are not relevantly similar to cases like The Server that concern us.

112 See Brewster and Mallison (2009) for a helpful overview and critical discussion of the relevant sociological literature.

113 See Niederle and Vesterlund (2010) for a helpful overview and evaluation of relevant empirical work.
3.2 The Content of Understanding

As we’ve seen, in the cases like The Server, various features of the socio-epistemic landscape put believers at risk of taking P as being causally grounded in something characteristically negative about G. P’s commonness among G is accurately explained, however, by some social injustice suffered by Gs, and not by some vice or lack of virtue characteristic of G as prominent negative historical narratives suggest. This clues us in to the content of the relevant sort of understanding in cases like The Server: it involves understanding why P is common among G – namely, Gs are victims of a social injustice. By extension, it involves understanding the nature of the relevant social injustice and its consequences.

3.3 The Effects of Understanding

The relevant sort of understanding inhibits believers from mistaking P as causally grounded in something characteristically negative about G. That’s because it provides believers with an obvious reason to think P isn’t grounded in something characteristically negative about G – that is, it makes it obviously unreasonable to think that P flows from something characteristically negative about G.114 For example, if Spencer understands that Black diners tip substantially less than average because of low incomes levels that trace back to racism, it will be obviously unreasonable for him to think it’s because of a lack of generosity characteristic of Black people.

So understanding in these cases steers believers away from negatively evaluating G and, derivatively J, in light of P, and invites them to simply take P’s commonness among G as a descriptive (nonevaluative) fact. To put it another way, understanding steers believers away from a negative version of the belief in question and...

114 It might be objected that the two explanations of P’s commonness among G at hand – some social injustice suffered by Gs or some vice or lack of virtue characteristic of G – are not mutually exclusive. Imagine some G who is oppressed and whose members cope by becoming callous or cruel or vicious in some other way so that, eventually, this vice becomes characteristic of G. In this case, P’s commonness among G may be accurately explained both by social injustices suffered by Gs and also by a vice characteristic of G. Such cases do show that the two explanations at hand are not mutually exclusive, but they are not relevantly similar to the cases that concern us. In cases like The Server – and, it seems, in most cases where the targeted social group is an expansive, culturally diverse social category like Black Americans or girls – social injustices cause P to be common among G, not by way of causing most Gs to become vicious or less virtuous, but rather by systemically disadvantaging Gs across social contexts –for example, by causing low levels of income, thwarting social mobility, hindering access to education and healthcare, and sapping G of political power and representation.
towards a nonevaluative version. That’s the first effect of understanding which makes it morally important.

The second effect is that it compels believers to recognize social injustices suffered by G.

### 3.4 Understanding Prevents Moral Bad

Because of its effects on believers, understanding prevents moral bad by reducing both the chance and the severity of harm risked by the beliefs in question. At a general level, the negative versions of these beliefs are more harmful than the nonevaluative versions. If they are negative, the beliefs in question pass unfavorable judgment on their objects, often unjustly. The mere possibility that they pass such judgment seems to be why beliefs like Spencer’s sting so much, why they pack a punch, why they hurt when we learn of one about us, why we bristle when we learn of one about another – it’s not unlikely that these beliefs say that there’s something characteristically negative about an already marginalized social group and, derivatively, its individual members.

To the extent that these beliefs risk harm because of the possibility that they pass unfavorable judgment on these individuals, understanding mitigates harm by inhibiting this possibility.

The relevant sort of understanding also stops the vicious feedback loop cultivated by the socio-epistemic landscape. When it is understood that the properties in question do not reflect poorly on G but, rather, trace back to social injustices suffered by G, the mistaken view is not confirmed and so does not go on to justify and perpetuate the social injustice.

What’s more, understanding seems to mitigate each of the specific harms that revisionists suggested were risked by the beliefs in question. One suggestion was that these beliefs offend the status of the relevant individuals when they do not have the property they are predicted to have. Another suggestion was that they violate the individuals’ interest in having beliefs about them be sensitive to how they actually are rather than how they are expected to be based on their apparent social groups. Both of these harms can be more and

---

115 It may be helpful to think of the negative version of the belief as having an implicit conjunct that captures the idea that P is causally grounded in something characteristically negative about G which J, as a G, likely instantiates: J is likely P and this reflects poorly on G and, derivatively, J.

116 I want to flag here that I have doubts about whether there is such a harm in general, though I am assuming that there is for the sake of discussion. Beliefs about individuals held by those with whom they have relationships are informed by individual-specific evidence available to them, and so this risk is posed by beliefs about individuals held by strangers. While it is plausible that individuals have an interest in their family and friends, for example, having beliefs about them that respond to how they actually are, I doubt that this interest extends to beliefs about them held by strangers; and, even supposing it does, I doubt whether there is a particular harm done to individuals just by virtue of strangers having beliefs
less severe, and they are less severe when the beliefs in question are nonevaluative: a belief that offends by mistakenly passing negative judgement on the character of its object is more harmful than a belief that offends by mistakenly attributing some nonevaluative property. Similarly, a belief that violates the relevant interest by insensitively passing negative judgment on its object is more harmful than a belief that is merely insensitive. For example, Spencer believing that Jamal’s being likely to tip less than average reflects poorly on Jamal as a Black person is more offensive, and exhibits a worse type of insensitivity, than Spencer believing that Jamal is statistically likely to tip less than average. So understanding mitigates both of these harms by steering believers away from the negative version of these beliefs and towards the nonevaluative version.

It was also suggested that these beliefs harm by leading the believer to act in ways that harm the individuals who are the objects of belief. Whether believers are less likely to act in harmful ways if the beliefs in question are embedded in the relevant sort of understanding is an empirical matter, and so all that can be offered on this point is speculation. That said, it seems plausible that believers will be less likely to act in harmful ways if they have the relevant sort of understanding. Suppose, for the purposes of this discussion, that Jamal is in Spencer’s section. If Spencer believes that Jamal is likely to tip less than average because of a lack of generosity, rather than a lack of information about tipping norms, he is more likely to give Jamal poor service – perhaps because there is a sense that miserly diners, but not uninformed diners, deserve or warrant poor service. What’s more, his understanding presents Spencer a lot of moral reason to not give Jamal poor service: since Black people have been systematically disadvantaged by racism – and, in fact, it’s ultimately because of this that Black diners tip less than average – Spencer can see that it’s obviously morally bad for him to act in ways that further disadvantage people like Jamal. So it seems plausible that understanding makes it less likely that the beliefs in question will lead the believers to act in harmful ways not only by steering them away from the negative version of these beliefs, but also by compelling believers to recognize the relevant social injustices.

The final suggestion is that the beliefs in question risk contributing to collective harms: they lead to harmful actions which, together, partly constitute collective harms. Moreover, these beliefs are a cog in the vicious about them based on information about their apparent social groups. Does the belief that he likely has diapers on hand really harm a parent out with his infant? And does the belief that she likely celebrates Easter really harm a Catholic?
feedback loop that perpetuates the social injustices that ultimately explain why P is common among G. We have seen how having the relevant sort of understanding interrupts the vicious cycle, and renders harmful actions less likely. By extension, when embedded in the relevant sort of understanding, the beliefs in question are less likely to contribute to collective harms.

3.5 Understanding Enables Moral Good

In addition to preventing moral bad, understanding enables, and even promotes, moral good. In order to organize and implement the sort of systematic social reform that is needed to remedy social injustices, believers must have the relevant sort of understanding. After all, it is impossible to organize and implement such reform without a firm and accurate grasp on what the social injustices are, why they are there, how they are perpetuated, and who is affected by them. Understanding, then, is necessary in order to edge society closer to justice. What’s more, understanding has the potential to motivate believers who have it to undertake such reform, thus promoting moral good, not just enabling it.

4. The Evidentialist Approach and The Informed Spencer

By appealing to the moral importance of understanding, the evidentialist approach, unlike the revisionist approach, can accommodate a plausible evaluation of variations like The Informed Server.

The evidentialist approach entails that Informed Spencer’s belief is rational and, thus, not prejudiced: it is properly based on Spencer’s statistical evidence. But just because evidentialism rules out Informed Spencer’s belief being prejudiced doesn’t mean it rules out its being morally condemnable: all prejudiced beliefs are morally condemnable but not all morally condemnable beliefs are prejudiced.

The evidentialist approach does entail that it’s not the case that Informed Spencer’s belief is morally condemnable because it fails to pass an evidential threshold set high by morality. But it does not entail a general verdict about whether Informed Spencer’s belief is morally condemnable. Evidentialism, however, is compatible with many theories about belief and morality that say that it not morally condemnable – perhaps because morality doesn’t place demands on belief or perhaps because there are moral demands on belief but Informed Spencer’s
belief satisfies by them. This verdict seems more plausible than the revisionist verdict that it is morally condemnable.

The evidentialist approach also has room to accommodate the thought that there may even be something morally commendable about Informed Spencer’s belief – at least, when it is taken as a part of a whole. Recall that one of the reasons that the relevant sort of understanding is morally important is because it enables, and even encourages, believers to organize and implement the systematic social reform that is needed to remedy social injustices. Such understanding should include beliefs about who is affected by the relevant social injustices – basic beliefs about the affected social groups, and also derivative beliefs about individual members of affected social groups. For example, in order for believers to recognize the need for and then to implement criminal justice reform, believers need not only basic beliefs about incarceration rates among Black men but also derivative beliefs about particular Black men. As Gardiner (2018) explains, “central to [social] injustice is the effect on individuals’ life chances… when a particular person is incarcerated, underemployed, participating in a crime, and so on, one potential source of injustice is their race, gender, or other social category means the outcomes was more likely. And these are social facts we ought to acknowledge.” It is only by having derivative beliefs about affected individuals that believers can grasp the full extent of the consequences of these social injustices, and realize the full extent of what needs to be done to address them. Beliefs like Informed Spencer’s are morally commendable, then, insofar as they play an important role in a morally good function of understanding. Contrast this with the revisionist approach whose stance on derivative beliefs prevents revisionists from condoning, much less commending, beliefs like Informed Spencer’s.

So, unlike the revisionist approach, the evidentialist approach does not entail the unintuitive verdicts that Informed Spencer’s belief is irrational, morally condemnable, and prejudiced; what’s more, it has room to commend such beliefs as morally important parts of the relevant sort of understanding. Even so, it may be argued, the revisionist approach at least has room to say something plausible about the moral significance of the harm risked by many derivative beliefs in question: morality requires “extra care and diligence” when it comes to beliefs about others that risk harm to them and, thus, morality does not condone a belief that risks

\[117\] Gardiner (2018), p. 182
such harm unless it has a lot of good evidence in its favor. If the evidentialist approach is unable to say anything along this front, then it may be worth accepting the revisionist alternative and biting the bullet on the unintuitive verdicts in cases like The Informed Server.

But the evidentialist approach does have room to say something here. In fact, evidentialists can affirm along with revisionists that morality requires “extra care and diligence” when it comes to believing things about others that risk harm to them. Here’s one thing that an evidentialist might say: the harm risked by beliefs like original Spencer’s gives believers like him strong moral reason, and perhaps even a moral obligation, to “do their homework” upon learning of Ps commonness among G, especially in cases where the moral stakes are high. For example, when a derivative belief risks the harm inherent to passing unfavorable judgement on G and, derivatively, J, morality might demand that believers double-check their evidence, review their reasoning process, and seek out information that could (dis)confirm it, contextualize it, or explain it. In cases like The Server, thoroughly “doing their homework” will lead believers to develop the sort of understanding found in The Informed Server which ends up significantly reducing the risk of harm.

So evidentialists and revisionists can largely agree about the moral significance of the harm risked by the beliefs in question, while disagreeing about what, exactly, the “extra care and diligence” required by morality in light of this risk amounts to: the revisionist approach says that morality requires believers to not have the beliefs in question if they lack of individual-specific evidence while the evidentialist approach can say that morality requires believers – or, at least, gives them reason – to act in ways that lead them to develop the sort of understanding which mitigates the harm risked by such beliefs risked. And, notably, it’s the evidentialist approach that seems to better delineate the “extra care and diligence” morality may require in light of the harm risked since it seems that Informed Spencer’s belief is morally in-the-clear given his understanding and despite his lack of individual-specific evidence.

Now let’s turn to what evidentialists can say about the Informed believers themselves. The evidentialist approach, of course, has room to say Informed Spencer is not morally condemnable: it does not entail that Informed Spencer has failed to meet some moral demand by dint of having his belief and it is compatible with many theories on which he isn’t morally condemnable. What’s more, the evidentialist approach has room to
accommodate the thought that there is something morally commendable about Informed Spencer when we evaluate him in light of his understanding by offering an alternative picture of “being woke.”

This picture may start with the thought that the revisionist picture leaves out what seem to be exemplars of “wokeness” – namely, believers like Informed Spencer who have the sort of understanding which enables them identify the presence and full effect of social injustices around them. On the revisionist picture, these believers are morally condemnable because of their derivative beliefs, and their understanding does not exonerate them. Evidentialists, however, can paint a picture of “being woke” that recognizes these believers as exemplars. “Being woke”, on this evidentialist picture, is a moral virtue that involves having a robust version of the relevant sort of understanding – which includes the derivative beliefs that make it fully effective – and, perhaps, being disposed to use it to help rectify the relevant social injustices. If something along these lines is correct, Informed Spencer may exemplify a moral virtue and, to that extent, should be commended in light of his understanding rather than condemned in light of his derivative belief.

5. Concluding Considerations

We started off with intuitive verdicts about cases like The Server – the beliefs in questions seem to be both prejudiced and properly based on supporting evidence – that revisionists argue not only undermine evidentialism but also motivate moral encroachment. In light of this, evidentialists are left with two tasks: to show that the relevant intuitions can be explained away and to show that these cases do not clearly motivate moral encroachment. This paper has focused on the second task, developing a challenge to revisionism by pointing to case variations in which the believers understand morally important features of their socio-epistemic landscape. The revisionist approach seems to entail the wrong verdicts, while the evidentialist approach is able to accommodate the right ones.

Of course, in order to show that, on balance, the cases at hand favor the evidentialist approach more than a revisionist approach, evidentialists still need to take on the first task. To do so, they must explain away the two intuitions about the original versions of cases like The Server that conflict on evidentialism. Given space constraints, this task must be left for another paper. To conclude, though, I want briefly gesture at a way that
evidentialists can approach this first task by appealing to the complex socio-epistemic landscape in these cases – a landscape ripe for muddling intuitions.

As we’ve seen in our exploration of the moral importance of understanding, the socio-epistemic landscape at hand puts believers at risk of accepting a mistaken explanation of why P is common among G and, thus, why J is likely P. Specifically, believers are at risk of taking P as causally grounded in something characteristically negative about G that J, in turn, likely instantiates. In the original version of cases like The Server, the possibility that the risk obtains is very much live. If the possibility of this risk does obtain, the evidentialist approach would categorize the doxastic state of the believer as prejudiced but also as epistemically irrational. After all, believing that P is grounded in something characteristically negative about G is not supported by the believer’s statistical evidence, and it’s also negative, and this is exactly the sort of belief that the evidentialist approach categorizes as prejudiced. But if the possibility of this risk does not obtain, and the believers believe that J is likely P without also accepting the mistaken explanation of why P is common among G, then the evidentialist approach would categorize it as neither irrational nor prejudiced. After all, such a belief would be supported and, thus, likely based on, the believers’ statistical evidence and so cannot be prejudiced.

Given all this, the evidentialist can say that, on the one hand, it’s the live possibility of the risk in question that triggers the intuition of prejudice when we consider the original cases. On the other hand, the intuition that these beliefs are properly based on the evidence is triggered by the fact that the belief that J is likely P is the only doxastic state explicitly ascribed to the believer. But, in fact, these intuitions aren’t about the same doxastic state. Rather, they are about two different possible doxastic states, only one of which obtains, but it is unclear which.

If something like this is right, then evidentialists can account for the initial intuitions about the beliefs in question in cases like The Server, and so do they not pose a challenge for evidentialism. It’s worth noting that this evidentialist account of the intuitions in the original cases is confirmed by intuitive responses to the Informed variations. When it’s clear that the possibility of the relevant risk doesn’t obtain – when it’s clear that the believers do not accept the mistaken explanation of why P is common among G – the beliefs seems to be both morally and epistemically in the clear.
In sum: revisionists and evidentialists can both agree prejudice shapes the world that gives believers evidence for beliefs. But, evidentialists add, the problem of prejudice taints true beliefs about prejudiced-shaped parts of the world almost exclusively when believers encounter these parts without understanding the relevant socio-epistemic landscape, and then go on to explain it in inaccurate ways that go beyond the evidence.\footnote{This claim is qualified with “almost exclusively” since it’s possible for a derivative belief to be based in part on internal or external prejudice rather than properly based on the supporting statistical evidence. Beliefs like these are accurate beliefs about a prejudiced-shaped part of the world that are, nevertheless, prejudiced and irrational even though they’re not accompanied by a mistaken explanation of why P is common among G.} The evidentialist approach seems to be the better approach given that, when the beliefs are embedded in the relevant sort of understanding, they seem to be unproblematic both morally and epistemically.
Works Cited


