

Benefiting from Epistemic Injustice

Ole Koksvik

Draft. Please do not cite without permission.

ABSTRACT

This paper argues that benefiting from epistemic injustice can generate stringent moral obligations with distinctively epistemic content, and that considering epistemic injustice allows us to better appreciate how benefiting from injustice, as a source of moral obligation, should be theorised more generally.

Two recent developments in political theory are the following. First, Miranda Fricker (2007) has argued that people suffer a distinctively epistemic form of injustice when they are harmed in their capacity as knowers. Second, a range of authors have joined Daniel Butt (2007) in arguing that benefiting from injustice ('BFI') can be a source of moral obligation. This paper's primary aim is to show that benefiting from epistemic injustice can give rise to moral obligations with specifically epistemic content.

If this is right, it is a significant result. Fricker argues that epistemic injustice is pervasive. Moreover, it is argued below that more cases should be recognised as instances of epistemic injustice (§1.2), and that people benefit from such injustice in a large proportion of the cases in which it occurs (§5). There is a large class of stringent moral obligations which has thus far been overlooked.

A secondary aim is to demonstrate the fruitfulness of bringing these two debates together (§4). The literature on BFI has been dominated by cases where the harm is practical (often monetary). Epistemic injustice involves harm of a different sort, and is therefore, as we shall see, especially useful for working out how BFI as a source of moral obligation should correctly be theorised.

My focus will be on the obligations a beneficiary of a particular injustice has to the victim of that injustice. However, in benefiting from injustice of any kind a person plausibly also acquires moral obligations to address the structural causes of those injustices. This is so in the epistemic case as well. While such obligations are important, they will not be addressed here. The obligations we have to one another are important too, and there is more than enough to say about them.

1. The State of Play

Appreciating the case for the Core Claim requires understanding Fricker's notion of testimonial injustice ('TI') in some detail.¹ Let's begin with a simple example.

Suppose Sally tells Holly that Arsenal won the game mainly due to the new striker's outstanding performance, and that Holly increases her credence in that proposition from .6 to .7 as a result. As a first pass, Sally suffered TI in this situation if Holly would have increased her credence *more* but for a prejudice she holds against people like Sally.

This gets close, but is not quite correct. To see why, note that when a speaker S says that *p* to a hearer H, this gives rise to a *rational challenge* for H, namely that of deciding how much more strongly she should believe that *p*, given this new evidence in its favour. Fricker notes two considerations of particular relevance to this challenge: H's assessment of S's *competence* with respect to *p* (does S know what she's talking about?), and H's assessment of S's *sincerity* about *p*, to hearers like H, in conditions like those prevailing (is S likely to tell me the truth about *p*, here, now?). These two assessments together constitute H's *credibility judgement* of S. There is, she argues, a simple, non-mysterious fact of the matter as to how that judgement should be made: it ought to reflect available evidence.² TI occurs when H's credibility judgement fails this test, and is instead depressed because of prejudice.

To see how prejudice enters the picture, return H's judgement of S's competence and sincerity. There are cases in which H has a lot of relevant information to help her: we know some people's areas of competence well, and we sometimes understand their incentives to tell the truth, or to lie. But often we know nothing about S's competence beyond what we know about the general level of competence in the population, and nothing about S's incentives beyond our knowledge of such incentives generally. Nevertheless, update our credence we must, since to not change one's credence in the face of new evidence is still to have taken a stance on that evidence.

Fricker argues that a rationally and morally unobjectionable response is for H to rely on *stereotypes*: widely held associations between social groups and one or more traits (p. 30). First, since we often lack specific information, doing so is *inevitable*. Second, many generalisations are *accurate*: that friendly locals

¹ In the book, Fricker discusses two forms of epistemic injustice: testimonial injustice, and 'hermeneutical injustice'. The latter is treated much more briefly than the former, and will not be discussed here. In what follows, unaccompanied page numbers refer to the book, unless context makes clear otherwise.

² "[T]here is no puzzle about the fair distribution of credibility, for credibility is a concept that wears its proper distribution on its sleeve. ... [T]he hearer's obligation is obvious: she must match the level of credibility she attributes to her interlocutor to the evidence that he is offering the truth" (p. 19).

are likely to give correct directions, and that family doctors are competent in diagnosing common ailments, for example. According to Fricker, the use of stereotypes to make credibility judgements is unobjectionable.

But not every instance of it is. Some stereotypes constitute *prejudices*, and when prejudices operate the result can be TI. According to Fricker a prejudice is:

Prejudice: An inaccurate stereotype [i.e., empirical generalisation] which persists despite (sufficient) counterevidence because of an ethically noxious affective investment.³

An example is the inaccurate stereotype that women cannot competently analyse sports, if it persists despite counter-evidence because of contempt against women.

We are now in a position to see how the first pass characterisation falls short. It said that S suffers TI if H increases her confidence less than she would otherwise have due to prejudice. But this is actually just a common but not inevitable *consequence* of TI. TI occurs at the very point when H assesses S's competence, sincerity, or both, less favourably than the evidence indicates—the credibility judgement is *unduly* deflated—because of prejudice. In so doing, H does S an injustice, and H is “*wronged specifically in her capacity as a knower*” (p. 20).

Testimonial Injustice: A hearer's credibility judgement of a speaker is unduly deflated because of prejudice.

When a credibility judgement is unduly deflated, TI has *ipso facto* occurred. This will often cause H to adjust her credence in the testified proposition less than she otherwise would have. But it need not, since other factors may cancel out the effect, or stop her from forming a belief at all.⁴ Whether credence is insufficiently adjusted or not—indeed whether *any* downstream consequence occurs or not—is immaterial to the question of whether TI has occurred. As we shall see, this is an important point.

People hold inaccurate generalisations about all sorts of social groupings, so TI can arise in a correspondingly wide variety of circumstances. However, some instances are more important than others. That is because some ways of placing people in social categories operate in restricted settings, while others have wide application. Suppose football players generally regard strikers as arrogant. This might affect strikers on the field and in related social settings, but it won't reach much further. Other social catego-

³ My interpretation here differs from that of Ishani Maitra (2010, 197), who thinks a stereotype being accurate is compatible with it counting as a prejudice on Fricker's account.

⁴ [Removed for blind review 1.]

ries operate in a wide range of settings. Examples include (perceived): race or ethnicity, religious affiliation, age, disability, body weight, gender, and sexuality. Such categories operate in what Fricker calls *tracker* prejudices; prejudices likely to track a person in a variety of different settings: “economic, educational, professional, sexual, legal, political, religious, and so on” (p. 27).

These categories are usually (always? p. 27) constitutive of a person’s *social identity*. When a person suffers TI linked to such a category, this usually goes along with a range of other injustices. TI which is connected with other injustices in this way is *systematic* (p. 27). Identity prejudices are also *persistent*: they change slowly if at all, so a person is often subjected to injustice their entire life. Accordingly, the central case of TI is *credibility deficit due to identity prejudice*.

I will now discuss some ways in which I think this theory needs amendment.⁵ The first is due to Ishani Maitra (2010). She argues that credibility judgements can be unduly deflated without this amounting to TI. To show this Maitra discusses Zara, who disregards a commentary article from a ‘tea-partier’ simply on the basis of prejudice against that group. Maitra argues that Zara commits no injustice because she “doesn’t owe it to [the author] to avoid a credibility deficit” (Maitra 2010, 199). We owe it to some but *not all* speakers to make correct credibility judgements.

Given the amount of information we are subjected to, and the scarcity of time and other resources, we cannot but rely on stereotypes, including, Maitra argues, the identity-prejudiced ones. Moreover, to stop ourselves would be unduly onerous, even were it possible (pp. 199-200). As we have “no *general* obligation” to make correct credibility judgements, accusations of TI must be backed up by an argument that in the case at hand such an obligation exists, or by theory to that effect.⁶

⁵ This is not an attempt at comprehensive review of the theory. There are aspects of it that I regard as implausible and / or insufficiently motivated, but which can be bracketed here. An example is Fricker’s insistence that prejudicial stereotypes operate “without doxastic mediation” (p. 36). The argument seems to reduce to the widely acknowledged point that the content of stereotypes may be inconsistent with a person’s beliefs. But this is easily accounted for in a doxastic framework, where the relevant conflict obtains between a person’s consciously held (or accessible) beliefs, and standing but not conscious (or not accessed / accessible) beliefs. There are also many engagements to which I think Fricker has convincingly replied, or which are not relevant here; see for example (Alcoff 2010), (S. Goldberg 2010), and (Hookway 2010), and Fricker’s reply (Fricker 2010).

⁶ In her forthcoming paper ‘Norms of Credibility’, Jennifer Lackey attributes to Maitra the view that a hearer owes it to a speaker to come to a correct credibility judgement only when the two already have a ‘special interpersonal relationship’, and then argues against that view (Lackey, Forthcoming). This misrepresents Maitra’s position. Maitra does hold that there can be testimonial injustice only when there’s an undischarged obligation to judge credibility correctly, and she thinks that special interpersonal relationships is one explanation for why such an obligation might exist. But there is no suggestion that this accounts for all cases. Quite the contrary: “The explanation I’ve just sketched is *by no means the only* possible one for why hearers are obligated to match their credibility judgments to the evidence in some cases of testimony, but not in others” (Maitra 2010, 201, emphasis added). Lackey’s critique therefore misses its target. In a separate point, Lackey convincingly argues that, while

The second amendment has the same origin. Fricker’s official formulation of a prejudice employs the notion of resistance to counter-evidence. Maitra convincingly argues that the more fundamental notion is a judgement “made or maintained *without proper regard to the evidence*” (p. 33, emphasis added), because one can fail to have proper regard to the evidence in other ways than by resisting counter-evidence (Maitra 2010, pp. 205–7).

A final amendment aims to allow the theory to accept support it intuitively gets from recent empirical research. There has been an explosion of work aimed at measuring *implicit bias* and its effects. Simplifying greatly we can think of implicit bias as an inaccurate negative stereotype a person is unaware of harbouring.⁷ For example, many will have observed that people who credibly profess to believe that males and females are equally capable in some domain nevertheless act contrariwise on occasion.

Fricker argues that TI is ubiquitous. (More on this in §5.) Empirical research has shown that implicit bias is likely widespread.⁸ It is natural to take the latter fact to support the former contention. Fricker’s

Maitra’s example of discounting a stranger’s testimony because she’s a tea-partier (to many) does not amount to testimonial injustice (Lackey does not share this view), this is not the case if you substitute ‘tea-partier’ with ‘woman’ or ‘black person’. But Maitra’s account is compatible with general and far-reaching obligations to historically oppressed groups that do not extend to groups without such a history, so this objection also fails.

⁷ With somewhat more precision, S is implicitly biased against members of a social group just in case some psychological fact about S of which S herself is unaware can predict and /or explain some aspect and /or degree of S’s negative actions or reactions, mental or otherwise, toward members of that group, which cannot be predicted and /or explained by reference to S’s consciously held attitudes with respect to that group. Note that this definition is non-committal on the nature of implicit bias but takes a stance on its explanatory power.

⁸ Some will regard this statement with scepticism or incredulity, because they know of recent research which casts doubt on the predictive power of the *Implicit Association Test*, which is a research paradigm on implicit bias (for example, Oswald et al. 2013). However, (1), as Alex Madva has argued, “[i]nferring from the fact that some IATs are bad predictors in some contexts to the general conclusion that IATs are bad predictors” is “pretty flawed reasoning” (Madva 2017), and, in any case, (2) rejecting the existence and causal efficacy of implicit bias on the basis of challenges facing *one particular research paradigm* is unwarranted. Implicit bias has been studied in a number of different ways, and its efficacy on behaviour has been established to a high degree of probability—indeed, “beyond reasonable doubt” (Jost et al. 2009). To take just one example, consider the seminal study by Steinfeldt, Anders, and Ritzke (1999), who sent identical copies of a CV—half marked with a fictitious male name, half with a fictitious female name—to a number of academics in psychology, asking for their evaluation of the candidate. The results were striking. Participants of both genders were significantly more likely to recommend hiring ‘male’ than ‘female’ applicants ($p < .001$), and significantly more likely to judge the applicant’s research experience, teaching experience, and service experience to be adequate (in all cases, $p < .005$) (pp. 520-21). These results indicate both the existence and causal efficacy of implicit bias—unless one is prepared to accept that a large fraction of the participants were *explicitly* biased against women.

Another important research paradigm is the *Affect Misattribution Procedure*. There a prime stimulus, a target stimulus, and a mask, are sequentially and briefly but supraliminally presented, before the participant is asked to rate the target stimulus. The target is ambiguous on the rated dimension; Chinese pictographs are (assumed to be) ambiguously ‘pleasant / unpleasant’

account does not admit of this support, since TI arises only when credibility is depressed due to *prejudice*, and many cases of implicit bias will fail to count as such.⁹ Fricker thinks a stereotype constitutes prejudice only if its resistance to counterevidence is explained by a *morally noxious motivation*, such as contempt for women (pp. 34–5). This is often implausible. For example, implicit preference for young people over old is often more plausibly explained by our culture’s constantly reinforcement of the ideal of youth, through advertising and other pop-cultural expression, and yet is surely often involved in TI.

I suggest that any case in which S suffers credibility deficit due to implicit bias, and where H is morally blameworthy either for harbouring the bias, or for allowing it to have practical effects, or both, shall count as a case of TI.¹⁰ This respects the spirit of Fricker’s account in that moral noxiousness remains a precondition of TI, but allows us to count many more cases of implicit bias causing credibility deficit as TI. Where implicit biases are widely known to be prevalent, people ought to act to limit the effects of this bias by their own hand.¹¹ Failure to do so will suffice for moral noxiousness. In this way Fricker’s claim that TI is widespread can accept the empirical support it intuitively gets from research on implicit bias.

for those who cannot read them, for example, and any residual difference is eliminated through aggregation. Statistical differences in how the target stimulus is rated can therefore be attributed to the unconscious influence the prime has on the rating task. For example, if a subject rates pictographs as significantly less pleasant when following a black person’s face than when following a white person’s face, implicit bias against black people is legitimately inferred. Clever tests have shown that that subjects are indeed most likely *misattributing* their reaction to the prime to the stimulus, that this effect is beyond and, importantly for present purposes, *often contrary to*, subjects’ conscious control, and that probably both affective and semantic information can be misattributed (thus rendering the paradigm’s name somewhat of a misnomer) (Payne and Lundberg 2014). This paradigm has e.g. demonstrated significant ($p < 0.001$) implicit bias against homosexual men / couples (Cooley, Payne, and Phillips 2013), black people, and the elderly ($p < .05$ in both cases) (Gawronski et al. 2010), and has been found to predict behaviour better, for example, than implicit association tests do (see Payne & Lundberg 2014, p.674).

Returning to IATs, for an introduction see for example Smith and Nosek (2010) or (Greenwald, McGhee, and Schwartz 1998). For results concerning *tracker* prejudices, for example race, gender, age, skin tone (dark / light), religious, ethnic, or cultural identity (Arab Muslim, Judaism, Native American), sexuality, and body weight, see e.g. (Nosek et al. 2007). For discussion of the test’s reliability, resistance to faking, construct validity, etc., see (Greenwald et al. 2009). For arguments that the states measured affect a range real-world behaviour see especially (Greenwald et al. 2009) for a large meta-analysis; and also (Payne 2001) and (Dovidio and Gaertner 2000). On the other hand, for an argument that the states tested by implicit association tests are not after all ones of which the agent is unaware—and thus not *implicit* biases—see (Hahn et al. 2014).

⁹ [Removed for blind review 11.]

¹⁰ I leave it open what the conditions for blameworthiness are; see (Holroyd 2012) for excellent discussion.

¹¹ Fricker has suggestions on how to do this; see especially her Chapter 4. However, many commentators doubt that individuals cultivating epistemic virtues will, by itself, be powerful enough to counteract the forces at play here; see for example (Langton 2010) and (Alcoff 2010). Interestingly, the results from the implicit association paradigm might lend Fricker qualified support. Given that “introspectively accessible attitudes and beliefs effectively guide deliberate actions but play weaker roles in determining spontaneous actions” (Greenwald et al. 2009, 20), the best advice may be exceedingly simple: make as large a proportion as possible of your socially relevant actions *deliberate*. But this is a topic for a different paper.

In what follows I assume that Fricker's account is amended as follows. First, there is TI only when a hearer fails to discharge an obligation to make a correct credibility judgement. Second, the more demanding notion of what 'proper regard for evidence' means is in play. Third, credibility deficit qualifies as TI in cases where the deflated credibility judgement is due to the operation of implicit bias for which the hearer bears moral responsibility.

1.3. *The Harm of Testimonial Injustice*

I will not take issue with Fricker's claim that when a credibility judgement is unduly deflated, S is *thereby* harmed in her capacity as a knower: my aim is to explore the consequences of accepting this view for the theory of BFI. On Fricker's account, two distinct types of harm occur in cases of TI. The primary harm is intrinsic; the secondary is extrinsic.

Many accounts of what gives human life its special value make reference to the human capacity for *reason*. To be capable of knowing, and of transmitting information, is on such views an important part of our humanity. To unduly deflate a credibility judgement is therefore, on Fricker's account, to *insult and dishonour* the speaker in a respect central to her full worth as a human being.¹² This injustice is *intrinsic*: if undue deflation of a credibility judgement occurs, on Fricker's view an injustice has *ipso facto* occurred.¹³

The secondary epistemic injustice relates to downstream consequences of the deflated credibility assessment. The central case of TI involves identity prejudice, which tends to be persistent and systematic. As a result of such treatment a person might lose confidence in her own beliefs. Since sufficient confidence is often taken as a requirement for belief, the person might end up with a diminished stock of knowledge.

¹² When a hearer unduly deflates her credibility assessment of a speaker, Fricker says that she in an important sense treats the speaker as *less than fully human*. Further, that the speaker is less than fully human is, on Fricker's account, (part of) the 'social meaning' communicated through the act (p. 44). I set these claims aside here, but those who accept them might find them relevant to the claims I make below.

¹³ This may seem to stretch the concept of injustice [Removed for blind review 2]. For suppose that the hearer hides her feelings well, and never acts on her deflated assessment. If the deflated assessment has no consequence, where's the injustice? We return to this question below (§3.1). Recall that the current task is to explore how the theory of epistemic injustice interacts with that of BFI as a source of moral obligation. In the latter literature the discussion is sometimes couched in terms of benefitting from *wrongdoing*—whether or not the wrongdoing constitutes an injustice (see e.g. Barry & Wiens 2014; Pasternak 2014; Lawford-smith 2016)—and the absence of consequences does not stretch the concept of wrongdoing in the same way. We are accustomed to thinking, for example, that one can wrong another person simply by thinking ill of her unjustifiably, or for harbouring ill feelings toward her for no good reason. In what follows I continue to speak of injustice, but readers should feel free to substitute with 'wrongdoing' as appropriate.

The person might also end up diminished as an epistemic agent. One of Fricker's examples is from *The Talented Mr Ripley*, in which Herbert Greenleaf dismisses his daughter-in-law Marge's suspicions about his son's disappearance because she is a woman. As this treatment persists Marge becomes what Greenleaf treats her as being: incoherent and hysterical. Less dramatically but still seriously, a person can fail to develop the intellectual virtue of *intellectual courage*: the ability to not give up too easily when one's views are being challenged. (Obviously, never giving up is not a virtue.)

1.4. Benefiting from Injustice

A person's actions can be a source of moral obligation. If I promise to do something for you I acquire a moral obligation to do it, unless stronger countervailing considerations present themselves later. Obligations also arise from *relationships*, from being a parent, for example. Slightly more controversially one can have moral obligations in virtue of one's abilities: if I can save a person from cardiac arrest at the cost of being late for dinner then I ought to do so. The question is whether merely benefiting from injustice, however innocently, is a source of moral obligation alongside these others.

To argue that it is, Daniel Butt (2007) describes an isolated island on which farmers A, B, C and D, live, separated from each other.¹⁴ Their yield depends on effort, and on water, which an underground river distributes evenly. Five hours of daily work ensures 200 kilos, which suffices for necessities. B, C, and D choose this course. A works all day, harvests 700 kilos, and enjoys the resultant abundance.

D sets out to redirect the river away from B and C toward herself, hoping to harvest more while working less. The plan fails; the water is redirected away from C and D, toward B. D kills herself in despair.

We're thus left with A, whose hard work yielded 700 kilos; B, who innocently benefited from D's unjust actions by harvesting 400 kilos from five hours' work; and C, who also worked five hours but now has nothing (Butt 2007, 132).

Someone should help C; the only question is *who* (Miller 2001). Taking an agent's actions, relationships, and abilities as the only sources of moral obligation means saying that the burden falls to A, in virtue of her greater ability. But intuition pulls in a different direction. C is the victim of an injustice of which B is the beneficiary. It seems that the obligation is on B to relinquish these benefits to save C.

Among those sceptical of BFI as a source of moral obligations the strategy has often been to argue that BFI goes along with something else. There is said to be a hidden element of contributing or upholding the wrong, for example, and this is said to be the real source of the obligation (Anwander 2005; Barry

¹⁴ Butt makes no claim to, and should not be thought of as having, come up with the idea that BFI is a source of moral obligation (though he did a lot to revitalise the issue). An important early paper is Thomson's 'Preferential Hiring' (Thomson 1973).

and Wiens 2014). I think this strategy fails. The island case rules out that a person's ability to help is the real source of the obligation. Luck egalitarianism has also been suggested as an alternative explanation (Huseby 2013), but cannot explain the directed nature of the obligations (Barry and Wiens 2014). And property rights also here. Even thinking individually about cases, then, things don't look good.

Stronger and more systematic replies are also available. First, the type of objection just considered cannot explain our sense that these cases *share* morally relevant features. Even if it were to succeed piecemeal it would be committed to the negation of this plausible claim. Second, strategy does not even get a foothold if we insist, as well we might, that the content of our intuitions—and therefore the beliefs we become justified in holding (omitted)—is not only *that* the obligations are acquired, but that it is *in virtue of benefiting from injustice* that they arise.

I think resistance to BFI as a source of obligation fails, but my aim is not to defend that claim, but to explore how epistemic injustice sheds light on how BFI should best be theorised. This question is independent of and strictly prior to the question of whether BFI should ultimately be regarded as a source of moral obligation. We must first identify the most plausible theory of it as such.

2. Moral Obligations with Epistemic Content

I will now argue for the paper's Core Claim:

Core Claim: Innocently benefiting from epistemic injustice can generate moral obligations with distinctively epistemic content.

Let's begin by clarifying what the claim means.

In the sentence "Given her evidence, S ought to have credence x in proposition p ", the 'ought' is *epistemic*; the sentence concerns the space of normativity S inhabits in virtue of being a *knower*. By contrast, the obligations in the Core Claim are not epistemic but *moral*, though they have distinctively epistemic content.¹⁵

¹⁵ The nature of the moral ought is of course deeply contested, but however we understand it, it comes apart from the epistemic one. For example, if I'm looking at a red ball on a table directly in front of me in good conditions, it is plausible that I epistemically ought to believe that there's a red ball in front of me, but not that I morally ought to believe that. Conversely, it is true that I morally ought to keep my promises but not that I epistemically ought to do so.

In saying that there are moral obligations with distinctively epistemic content I do not mean to claim that there are true claims of the form, ‘morally, S ought to believe that p ’. Fricker is reasonably interpreted as being committed to such claims, and related claims have been discussed by others.¹⁶ Perhaps such claims can even be supported by things I argue here.¹⁷ But I do not intend this meaning here.

Instead I have in mind obligations to perform actions *which address a person specifically in their capacity as a knower*. If I help you understand a logical truth I have performed such an action, and likewise if I direct my children’s attention to epistemically significant parts of a scene—the structure of an ant hill, say. When unpacked, then, the Core Claim says that *innocently benefiting from epistemic injustice can generate moral obligations to perform actions that specifically address a person in their capacity as a knower*.

I shall argue for the Core Claim by examining the following case:

Two Accountants. A large company is recruiting one junior and one senior accountant through a single process. The positions differ in pay, and time for professional development. After an initial, fair round, two applicants are left. They are equally well qualified in all respects, with the possible exception of experience, which is now up for consideration.

The only evidence for the candidates’ experience is their own accounts thereof, both of which are true. The candidates have different employment histories, but by the lights of the company’s guidelines, the histories are equally valuable. If both accounts are believed, the candidates will therefore be (correctly) assessed as equal in this respect also. According to company policy the decision is then to be made by the toss of a fair coin.

The committee harbours substantial uncertainty about what credence to assign the candidates’ accounts. It decides to make a determination after face-to-face interviews.

The candidates are visibly different in traits relevant to perceived ethnic identity. As a result of identity-prejudice on part of the committee members Vic receives a significant credibility deficit. His account of the depth of his experience is consequently given relatively low credence. His competitor Ben’s account is given higher credence. For this reason, Vic is offered the junior position, and Ben the senior one. That prejudice played this role later becomes clear to both, but there is nothing either person can do to reverse the process, start it

¹⁶ See e.g. (Stroud 2006), who raises the prospect of such obligations in the context of friendships, and (Goldberg, Forthcoming), who argues that reasonable expectations ground claims about what subjects *should have known*.

¹⁷ [Removed for blind review 3.]

anew, or to affect wider change within the company.¹⁸ Moreover, it is clear that the committee is not disposed to hear any complaints or take any action whatsoever.

By coincidence of timing, shortly after being offered the junior position, Vic is also offered, and takes up, a position in middle management. He is indifferent between management and accounting as career paths, the management position offers the same wage as the senior accounting role, and there are equivalent opportunities for professional development.

In *Two Accountants*, Vic was the victim of TI. As a result, an innocent person, Ben, benefited: his chances at the job he wanted went from 50 / 50 to a sure thing. What are the consequences for what they owe to each other?

On the account adopted in this paper (§1.3), we can be sure that Vic suffered a primary epistemic harm. When the members of the hiring committee unduly deflate their credibility judgements of Vic, the primary epistemic harm obtains *ipso facto*. The credibility deficit constitutes an intrinsic injustice against him: an insult and a dishonour to him as a knower. He also most likely suffered a secondary epistemic harm. Let's assume that Vic regularly suffers identity-prejudicial credibility deficits: it is a persistent and systematic feature of his life. As a result he has lost confidence in some propositions, and we may assume that he has lost some knowledge. Moreover, his intellectual development has been stunted; he has failed to sufficiently develop the epistemic virtue of intellectual courage.

What, if anything, does Ben owe Vic? Clearly he ought to act with sensitivity towards Vic. Losing out on a job is discouraging, and just like everyone else, Ben ought to be mindful of that. Also just like everyone else, he of course has a moral obligation to not insult and dishonour Vic further. Is that all?

I suggest not. When Ben becomes apprised of the situation he ought to *make it his business* to counteract the injustice from which he benefitted by attempting to counteract the insult and dishonour Vic suffered in his capacity as a knower. That entails taking actions that specifically address Vic in his capacity as a knower because to do so Ben must specifically honour him in that very capacity.

Ben might, for example, make an extra effort to listen especially well whenever Vic speaks. Listening well is not trivial. It takes effort. By making an effort that in normal circumstances would be supererogatory,

¹⁸ This assumption is merely simplifying, and nothing in what follows relies on it. That is because having the type of obligations which are my focus does not exclude having other obligations, for example to try to effect structural change over time. (See also §3.2, below.) It is not an implausible assumption, however. It is not uncommon for the stratum of people who makes decisions to be so radically separated from the rank and file of employees that there is very little influence from the latter group to the former—think of high powered law firms, for example.

Ben is honouring Vic in his capacity as a knower. Ben might also seek out Vic's view in particular, in situations in which there are others he might equally well ask, thereby making it especially clear that he values Ben's opinion. He might go the extra mile to support some of Ben's epistemic endeavours. And there are other similar options. It is not plausible that there is any particular action among these that Ben ought to perform. It is plausible that he ought to take some such action.

Furthermore, Ben ought to counteract further epistemically disrespectful behaviour towards Vic. Suppose that Ben, Vic, and a few others are gathered around the watercooler. One of the others repeatedly dishonours Vic in his capacity as a knower by interrupting him, interpreting him uncharitably, and not giving him epistemic credit where it is due. Ben ought to counteract this behaviour by calling out the interruptions, pointing out more charitable interpretations, and explicitly giving Vic credit. Here Ben cannot pick and choose: he ought to take *all* of these steps.

So Ben ought both to take active steps to honour Vic in his capacity as a knower, and to resist further insult and dishonour being perpetrated on him. Finally, he ought to publically condemn the injustice from which he benefited: he ought to publically state his disapproval for the committee's behaviour.

As for the secondary epistemic harms, Ben ought to take steps to counteract the loss of knowledge and stunted intellectual development Vic suffers as a result of his continuing exposure to deflated credibility judgements. However, since the harm Vic has suffered here was caused not only by the injustice from which Ben benefited, but from a long series of other injustices besides, exactly how much Ben in particular ought to do is less clear.¹⁹

In response to Two Accountants, intuition says (i) that Ben has these obligations, and (ii) that he has them *because* he benefited from the injustice suffered by Vic. If (i) and (ii) are accepted, the Core Claim is established. That claim, recall, says that innocently benefitting from epistemic injustice can generate moral obligations which specifically address a person in their capacity as a knower. Ben's obligations obviously qualify. If he has them because he benefited from epistemic injustice the Core Claim follows, by existential generalisation.

Moreover, we can see *why* and in which circumstances these moral obligations arise. Innocently benefitting from epistemic injustice generates moral obligations to take actions which specifically address a per-

¹⁹ A reasonable first stab might be that he must make a contribution proportionate to the damage done by the harm from which he benefited relative to the total secondary epistemic harm. However, the real answer probably depends, among other things, on the extent to which Ben is a member of a social group the members of which routinely benefit from injustices suffered by the group to which Vic belongs (Thomson 1973), as is, for example, the case with white and black people in the US.

son in their capacity as a knower *whenever and (in part) because* a person reasonably believes that such actions will most efficiently further the goal of counteracting the epistemic injustice from which he benefited. Given that (I shall assume) this is a widely, albeit implicitly, held belief about human psychology, benefitting from epistemic injustice generates moral obligations with specifically epistemic content most of the time.²⁰ So, (i) Ben's moral obligations are well explained by (ii) the fact that he benefited from the injustice Vic suffered, on plausible empirical assumptions.

Showing this is only half the job done. Opponents of BFI as a source of moral obligation have, as noted, tried to explain away our verdicts by offering alternative explanations of them. In this case, too, we must show that this explanation is better than competing ones. To this end I now consider a number of alternative explanations, and argue that (ii) (with the reasonable beliefs) is better than all of them.

Some of the alternative explanations invoked elsewhere are non-starters in the case of TI. For example, it has been suggested that some of the intuitions proponents of BFI as a source of moral obligation appeal to can be explained by a concern for property rights (Huseby 2013). But it is obvious that property rights cannot explain Ben's obligations to honour Vic in his capacity as a knower, to resist further insult and dishonour being perpetrated on him, or to publically condemn the injustice.

Similarly, we can set aside the alternative explanation that Ben's obligations can be explained by luck egalitarian considerations. First, Luck Egalitarianism is a theory of distributive justice, and (*pace* Coady (2017), credibility isn't a distributive good.²¹ Second, Luck Egalitarianism can at most deliver *undirected*

²⁰ The exception is when someone reasonably believes that some other type of action will most efficiently serve to counteract the injustice. We can imagine (strange) circumstances in which this would be so. Benefitting from epistemic injustice thus generates moral obligations with epistemic content only contingently. [Removed for blind review 4.]

²¹ As Fricker notes (2010). A correct credibility judgement properly reflects the available evidence about the speaker's competence and sincerity. Nothing stops me from making a correct judgement of this kind of both S and of S*, even when S asserts that *p* and S* that *not-p*. There is only so much *credence* to go around between the propositions *p* and *not-p*—if one is to remain rational it is usually thought that they must not add up to more than 1—but credence is not credibility. I can easily judge *both* S and S* to be highly competent and sincere, because I take them to have access to different evidence pools, or to evaluate evidence on the basis of distinct but (both) plausible theoretical paradigms, for example. Conversely, one can fail to (significantly) increase one's credence that *p* without unduly deflating one's credibility judgement of S even when S has just said that *p* because one has other evidence bearing on *p*. A different speaker having just asserted that *not-p* is one type of such evidence: the correct updating function on [S said that *p*] is often distinct from the correct updating function on [S said that *p* and S* said that *not-p*]. It might even be that the fact that someone equally competent as S utters that *not-p* indicates that I *should* deflate my credibility judgement of S, specifically the part of it that regards S's competence with respect to *p* (Worsnip 2014) argues for something like this view). Then the credibility judgement is *duly* deflated. But even if not, there is no reason to think that the reduction in credibility occurs because there is some resource to be distributed. A big increase in one's credibility judgement of

obligations: general obligations to anyone who has suffered bad luck. Ben's obligation is specifically to Vic. Luck Egalitarianism can't explain the directed nature of obligations from BFI.

This also shows that Ben's obligations do not stem from his *ability* to help. Obligations arising from a person's ability to assist are also undirected: they hold between the agent with the ability and anyone she is able to help. Consider the view that affluent people are morally obliged to ameliorate the suffering of the very poor in virtue of our ability to do so (Singer 1972). This is not an obligation to a particular poor person, but an obligation to anyone (equally) needy which we are (equally) able to help. If Ben's obligation to Vic were grounded in his ability to help him, the same would hold here.

It does not. Suppose Two Accountants is not unique; there are other cases like it, and Ben learns of some of them. *Ex hypothesi*, Ben then has a general, undirected obligation to all these victims of injustice. This entails that, if a situation arises in which counteracting the injustice suffered by one of the victims precludes counteracting the injustice suffered by a different one, it is, other things equal, morally arbitrary which one he chooses. That prediction is wrong. If a such a situation arises and Vic is one of the two victims, then, other things equal, Ben ought to counteract the injustice suffered by Vic. That Ben's obligation stems from BFI can explain this, since Ben benefited from that injustice, and not the others. That it stems from his ability cannot, for then he should be free to combat the injustice suffered by any other victim to the exclusion of Vic. Similarly, even on the assumption that other employees in the company are equally able to take steps specifically to honour Vic in his capacity as a knower, it is not plausible that these others have an obligation to do so. Ben, however, does.

This alternative hypothesis also fails to explain the strength of Ben's obligation to Vic. Suppose again that there are several other similar cases of injustice, and that Ben becomes apprised of them. His obligation to counteract the injustice suffered by Vic is stronger than his obligation to combat and counteract any of the other injustices, other things equal.

A good way to operationalise differences in strength between moral obligations is by reference to the cost people must accept in order to discharge them (Narveson 2003; Haydar 2003). For example, a person cannot easily justify violating a moral obligation to not seriously harm other innocent people by reference to the cost this imposes on her—she has to avoid this even if it costs her a once-in-a-lifetime employment opportunity, for example—and the costs she has to take on in order to rectify the situation if she does harm someone are significant (Barry and Wiens 2014, 21–22).

S* may necessitate only a small reduction in the credibility judgement of S, or, as we saw, none at all. Testimonial injustice is not a species of, and not usefully thought of on the model of, distributive injustice.

Counteracting epistemic injustice plausibly entails costs for Ben. He may suffer damage to his reputation and prospects for advancement, be more at risk of demotion as a result of minor mistakes, and so on. At some point such costs justify taking no action to combat the injustice suffered by one of the other victims. Likewise, at some point such costs justify taking no action to counteract the injustice suffered by Vic. However, the latter cost is greater than the former. Ben is not obliged to suffer just any consequences to speak up on behalf of Vic, but he is obliged to suffer more strongly adverse consequences on behalf of Vic than on behalf of other victims. Ben's obligation to Vic is stronger than to the other victims. As before, this is well explained by the hypothesis that it stems from BFI, since he benefited from the injustice suffered by Vic, but not from any of the other injustices. The alternative hypotheses cannot explain this.

A different explanation says that all people have a general obligation to fight injustice whenever we encounter it, and claims that this is the source of Ben's obligations.²² While it is plausible that there is such an obligation this cannot be the explanation here, for this view, too, fails to account for the directed nature of Ben's obligations, and for their strength. If Vic and another person are both subjected to epistemically disrespectful behaviour in Ben's presence, Ben ought to prioritise speaking up for Vic if he is forced to choose. And, while all who are present at the watercooler plausibly have an objection to speak up simply in virtue of encountering the injustice, Ben's obligation is stronger than the others' obligation in that he must accept a higher cost to discharge it. None of this can be explained by a general obligation to fight injustice whenever one encounters it.

Yet another hypothesis holds that Ben's obligations arise because he is a member of a class of people who routinely benefit from injustice towards a group of which Vic is a member. However, there is nothing in the specification of Two Accountants dictating that group memberships are as the objection needs them to be. The case specifies that the candidates are visibly different in traits relevant to perceived ethnic identity, but not what some may have assumed, namely that exactly one of the applicants is white. Both applicants may be African Americans, for example, but differ in degree of similarity to the stereotype of African American people. There is strong empirical evidence that prejudice against African Americans is stronger against those who are more similar to the stereotype (and evidence of the same phenomenon exists for other groups also, for example for Hispanics).²³

²² [Removed for blind review 5.]

²³ For an overview, see (Maddox 2004). For evidence of greater discernment of differences within one's in-group than outside of it, so-called 'out-group homogeneity', perhaps suggesting more skin tone bias against black people among black people than among white people, see (Linville, Fischer, and Salovey 1989). However, see (Maddox and Gray 2002) for evidence that both groups use skin-tone to socially categorize people they observe—skin tone is not, after all, masked by out-group homogeneity, in other words—and for evidence of the association of dark skin with negative traits and the converse: "Generally, dark skin tone

This opens the possibility that *both* candidates suffered identity-prejudicial credibility deficit (and thus TI) in Two Accountants, but Vic more severely than Ben. This does not change the verdict. As a beneficiary of injustice, albeit of a type of injustice he himself also suffered to a lesser degree, Ben still acquires the epistemic obligations outlined above.²⁴ He still ought to counteract the primary and secondary epistemic harms Vic has suffered to the best of his ability, and he ought to take on a higher cost than he would have had to were he merely a bystander. So Ben acquires epistemic obligations by benefiting from injustice even when he does not benefit but is instead harmed epistemically, so long as he benefits somehow.²⁵

In fact, the objection fails *even if* the group memberships are as it needs them to be. It is plausible that any member of a class of people who systematically benefit from injustice thereby acquires moral obligations (Thomson 1973), but Ben's being a member of such a class cannot explain his obligations in this case. Obligations arising from group membership are plausibly construed as obligations either a) to counteract such injustices whenever encountered; or b) to as efficiently as possible combat the system which enables these injustices; or, most plausibly, c) both. If Ben is aware of several instances of injustice, he ought to prioritise counteracting the injustice suffered by Vic over the other injustices, even if the victims are all members of the same social class. He has a directed obligation to Vic as an individual, and not merely an obligation, directed or otherwise, to the class of people to which Vic belongs (although he may have that, too). Moreover, even if Ben's counteracting the injustice suffered by Vic is not the most effective way to counteract the system, other things equal that is still what he ought to do.²⁶ As before, this hypothesis fails to explain features of the situation which the favoured hypothesis does explain.

One can also imagine someone claiming that anyone to whom the situation becomes transparent and who is also involved at least by being a bystander acquires the kind of directed obligation Ben has. One needs to be involved in the situation to obtain benefiting-sourced directed duties, but one need not be a beneficiary (nor a perpetrator); being a bystander is enough. But accepting this 'involvement-condition' on directed duties does not imply that the moral obligations of bystanders are on a par with those of

was more often associated with poverty, aggressiveness, lack of intelligence, lack of education, and unattractiveness" (p. 258).

For evidence from the (as noted, recently controversial; see n. 8 above) Implicit Association Tests, see e.g. (Nosek et al. 2007).

²⁴ Naturally, if Ben has suffered epistemic injustice, other agents may have acquired epistemic obligations to him.

²⁵ This is compatible with the conclusion that, in the watercooler scenario, the *overall* result is now that others should speak up instead. That's because, even though beneficiaries' obligations are stronger than bystanders', if Ben is himself an African American it will likely be more costly for him to speak up, possibly so much so that this outweighs the difference in strength of obligation. This is also a place where the obligations discussed in this paper interestingly interact with those which arise due to group membership [removed for blind review 6], but as before I must leave this to one side.

²⁶ This does not commit me to the implausible claim that *no* discrepancy in efficiency would dictate a different course of action.

beneficiaries. As noted, they differ in strength. Bystanders, such as the others at the watercooler, can appeal to quite small costs to justify not intervening. A beneficiary must appeal to higher costs.

The only even remotely credible explanation for why Ben has the moral obligations to Vic he does have is that he benefited from the TI Vic suffered. Attempts to explain his obligations in other ways fail. We thus have strong reason to believe (i) that Ben has these obligations, and (ii) that he has them because he benefited from epistemic injustice. The Core Claim is established.

3. Objections and Replies

3.1. *No Harm, No Foul*

We can imagine a person suffering TI but no further consequences. This might give rise to the objection that, contra Fricker, such a person is not really *harmed*, though she is *wronged*. When someone is wronged but not harmed, the objection continues, beneficiaries do not acquire obligations to victims of injustice. Consequently, to evaluate obligations from BFI we must consider what would have happened if no injustice had taken place. We must evaluate two variations of the Two Accountants case. On the first, if the committee had not been prejudiced it would have hired Ben anyway. On the second it would then have hired Vic. In the former case Vic was not harmed, for he suffered no adverse consequence relative to a just state of affairs. He is therefore owed no compensation, by Ben or anyone else. In the latter, Vic is owed compensation, but that is because the decision was faulty, and not because Ben benefited from the injustice. So in a portion of the cases Vic is owed nothing, and in the cases in which he is owed something, an alternative explanation can be provided.²⁷

This objection fails. First, this paper concerns the consequences of *accepting* Fricker's account of TI for the theory of BFI. It is no objection that they might not follow if we reject it.

Second, the objection trades on the distinction between benefiting from injustice *per se*, and benefiting from a nearby downstream consequence of injustice, a distinction of doubtful import in this context. In Butt's island case, for example, the real injustice is *that C is deprived of water*. That would be unjust even if the fresh water were uselessly directed to sea. Thus, even in this paradigm case B in *some* sense benefits from a downstream consequence of injustice, rather than from the injustice itself, and yet we do not hesitate to say that B acquires moral obligations as a beneficiary of injustice. By parity of reasoning, even if

²⁷ This objection is thus a variant of the 'explaining away' strategy considered in §2 above. [Removed for blind review 7.]

Ben *in some sense* benefits from a downstream consequence of injustice rather than from injustice itself, we should not hesitate to say that he acquires moral obligations as a beneficiary of epistemic injustice.²⁸

Third, while there may be some cases in which we know what would have happened absent injustice, in practice this is often not the case. People in Ben's and Vic's positions certainly do not usually know, nor do the members of the search committee. The latter will often be unaware of having been influenced by prejudice in the first place,²⁹ and unlikely to be able to reach a justified conclusion about how they *would* have acted otherwise, were we to ask.

Fourth, it is widely accepted not only i) that what would have happened absent injustice is often not knowable, but ii) that one need not know this in order to assess obligations from BFI. For example, one need not know whether a victim is net worse off as a result of the injustice to evaluate her claim: after all, the injustice might have fortuitously saved her from much greater calamity. Instead it suffices to know that one of her legitimate interests is illegitimately set back by the injustice (Butt 2009, 109; Woodward 1986, 809; Fabre 2007, 143; Barry and Kirby 2015, 6–7). Vic has a legitimate interest in his testimony being assessed without the influence of prejudice, an interest which is illegitimately set back.

But most importantly, in *Two Accountants* it is not merely *difficult to know* what would have happened absent injustice, *there is no fact of the matter*. Absent injustice the outcome would have been decided by a coin toss, one which, as things actually turned out, never took place. Nevertheless, Vic clearly suffered epistemic injustice, and Ben clearly benefitted from it. The committee's credibility judgement of Vic was unduly deflated due to prejudice: that's TI. This caused Ben's chance at the superior job to go from 50/50 to a sure thing. To deny that this is to benefit would be absurd.

3.2. *Obligations, But Not To the Victim*

It has been objected that BFI does not give rise to obligations directly to victims, but instead to obligations to change the sources of the injustice: the structural conditions that give rise to injustice in the first place. Holly Lawford-Smith has argued, in the context of benefiting from class privilege, that the beneficiary making herself worse off would (often? usually?) benefit nobody, at least not in a just way, and moreover that it "might not" improve the future with respect to just distribution of goods (2016, 40–41). Similarly, in the context of racial discrimination, Jeremy Dunham and Lawford-Smith (2017, 17)

²⁸ Some might say that Ben's obligations are not aptly characterised as duties to *compensate* Vic. No matter, give them the term. The Core Claim says that innocently benefiting from epistemic injustice can generate moral obligations with specifically epistemic content. Ben's obligations obviously qualify, so how we label them is not important. See also n. 31 below.

²⁹ This underscores the importance of allowing testimonial injustice to encompass cases of implicit bias, as urged in §1.2 above.

argue that “[w]hat is owed should be directed at the sources of the morally problematic race-based differences, not to .. the disadvantaged”, because “[d]irecting benefits towards the race-disadvantaged directly ... might perpetuate the status quo [by], creat[ing] ad hoc redistributions ... without upsetting the actual systems and institutions” that generate the injustice in the first place. An analogous objection could be raised against the proposed account of Ben’s obligations: when he benefits from the injustice Vic suffers Ben does not acquire obligations *to Vic*, but instead to try to ‘upset’ the systems and institutions which give rise to the injustice in the first place.

In reply; first, though beneficiaries of epistemic injustice also acquire obligations to try to change the structures that generate injustice, this does not show that they do not also acquire obligations directly to the victims. Second, in the cases at issue in this paper (and others), the recommended actions certainly *would* benefit the victim in a just way. And third, as I have argued elsewhere (omitted), perversion of incentives is in general a weak objection to a moral principle, because it focuses on only one part of what should motivate the relevant actors. Of course many actors fail to be motivated by things that should motivate them. What of it? The mere possibility of a situation in which an agent, cognisant of the moral principle under discussion but ignoring other important principles, fails to be correctly incentivised, is no objection, for such situations can be generated for any (non-total) moral principle, and so would not leave any left standing. Dunham and Lawford-Smith seem to suggest that ad-hoc redistributions remove an incentive those who sustain the status quo would otherwise have to reform it, namely the bad consequences for victims. Perhaps. But what *should* concern those people is not the position disadvantaged persons end up in after costly interventions, but where they would have ended up without them. That some actors fail to take this into account and may thereby be wrongly incentivised says little about the truth or falsity of moral principles.

4. Lessons from the Epistemic Case

This paper’s secondary aim is to demonstrate the usefulness of considering epistemic injustice for theorising BFI as a source of moral obligation. The discussion has already begun to bring this out: we saw that some alternative explanations don’t get off the ground in the case of epistemic injustice. I will now outline some of the further theoretical lessons I think we can learn by considering the epistemic case.

4.1. Beneficiaries’ Obligations Are Not Derivative

In *Two Accountants*, Ben’s obligations pertain to the intrinsic injustice that obtains simply in virtue of Vic suffering a credibility deficit. Reflection on this feature of the case is particularly illuminating with

respect to BFI more generally. In the literature it is sometimes held that BFI gives rise to moral obligations only insofar as, and at least in part *because*, the perpetrator of the injustice is not there to shoulder it. The obligation to rectify the injustice is thought to fall primarily on the perpetrator, and only secondarily and derivatively on the beneficiary.³⁰ A corollary is that, in the cases of practical harm which have so far dominated the literature, beneficiaries' obligations can cease to exist if someone else—the perpetrator, a collective insurance scheme, or what have you—fully compensates the victim.

Cases of benefiting from epistemic injustice show that beneficiaries' obligations are not *in general* derivative. We have been misled to think so by the facts that a) beneficiaries' obligations cease if the victim is 'made whole' again (Goodin and Barry 2014); b) in the practical / monetary case, being made whole again usually *just means* being fully compensated, since money and (generally) property are fungibles; and c) the perpetrator's obligation is stronger than the beneficiary's.

Suppose one of Alice's most treasured possessions is a cigar which Alice's father, who recently succumbed to cancer, bought on the day she enrolled at university, hoping to live to smoke it at her graduation. Alice's friend Bob doesn't know this. While visiting Alice he realises that he's forgotten to get his husband Chad an anniversary gift. Chad loves cigars. Planning to replace it the next day Bob takes the cigar, wraps it up, and gives it to Chad. Chad appreciates the gift, and smokes it.

Alice suffers a practical harm here, but she can easily be made whole again in that dimension. She also suffers an emotional harm. She simply cannot be made whole again there, at least not in the short term. Nevertheless, Bob *and* Chad are obliged to try as best they can: they ought to do whatever they reasonably think has the best chance of making her whole again emotionally. There is nothing derivative about Chad's obligation here. Moreover, Bob and Chad really do have a moral *obligation* to do this (and not just a reason): both would appropriately be criticised if they do not. Chad has a stronger obligation than bystanders do, though not as strong as Bob.

There are many cases of emotional harm where the victim of the injustice can be made whole again—apologies often work. But in at least some such cases, and in at least some cases of epistemic injustice, there seems to be nothing the perpetrator can do which will relieve the beneficiary of her obligation. Plausibly, the only way Vic can be made whole again is if *everyone* involved in the injustice—perpetrators *and* beneficiaries (and maybe also bystanders)—publically repudiate the dishonour, for example by explicitly condemning it, and by honouring the victim in his capacity as a knower. This difference between the epistemic and the practical cases traces to the fact that the primary harm is intrinsic—a feature epistemic harms share with emotional ones.

³⁰ See e.g. (Huseby 2013, 3). [Removed for blind review 8.]

Reflection on the cases of benefiting from epistemic injustice thus leads to a more general insight about BFI: while beneficiaries' obligations may be secondary and derivative in cases of extrinsic harm they are not where intrinsic harm is concerned. Indeed, although closer investigation is warranted, the simpler and thus preferable theory says that beneficiaries' obligations are *never* derivative or secondary, although they are weaker than those of perpetrators.

4.2. Still No Demarcation Criterion

Let's turn to accounts of the demarcation criterion between cases of BFI that do, and those that do not, give rise to moral obligations. Haydar and Øverland (2014) suggest that at least one of three further conditions have to be present: a structured competition that is perverted by the injustice, that the perpetrator be motivated by a desire to benefit the beneficiary, or that assets be transferred to the beneficiary.

The epistemic case shows that these cannot be disjunctively necessary conditions for BFI to generate moral obligations. Two Accountants fits, since the epistemic injustice perverts a structured and otherwise fair competition, but there are cases of epistemic injustice that do not.

Consider three friends discussing the likely outcome of an upcoming election. If B and C disagree about a point where judgement comes in, and A gives B's claim more credence than C's because of identity prejudice (e.g. against her gender), the first condition is not met. There is also no transfer of assets, nor is there a reason to think that A did this with the intention of benefiting B. Yet we can flesh out the case so that B benefits, and, as a result, B obtains more stringent obligations to counteract the harm which C has suffered than does a mere bystander (D, who is not participating). Having formed the belief that B 'won' the discussion A might be more inclined to believe B in the future, B's social status may rise in A's eyes, or, more mundanely, A might afford B material benefits which he does not afford C, as a reward.

According to Barry and Wiens (2014), BFI gives rise to obligations only when the beneficiary's retention of the benefit *sustains the harm* suffered by the victim.³¹ On their account, there is a tight and substantial link between the two:

³¹ It is not entirely clear whether these authors intend the kind of obligation I have been discussing to fall within the purview of their account, but if they do, epistemic injustice is a counterexample. For example, they say that the account is not intended to encompass obligations to apologise. On the other hand, they accept as within the purview of the account cases in which "materially benefiting from wrongdoing triggers special duties to the victims of that wrongdoing" (p. 5), and, as I have argued, Two Accountants is just such a case. Their restriction to cases of remedial duties is also met: the question is precisely "whether those who innocently benefit from the wrongdoings in question thereby acquire duties to ... *act on [victims'] behalf* to rectify the wrongful state" (p. 5, emphasis added).

An innocent beneficiary of wrongdoing owes a benefiting-related remedial duty to the victim(s) of the wrongdoing if and only if the beneficiary's receipt and subsequent retention of the benefits from the wrongdoing sustains wrongful harm (p. 11).

It's the retention that *sustains* the wrongful harm.

[B]eneficiaries incur a benefiting-related remedial duty to the victims of wrongdoing if and only if they sustain wrongful harm *by* retaining the benefits (p. 18, emphasis added).

In Two Accountants, however, there is no sense in which Ben's retention of the benefit sustains the dishonour to Vic, and yet Ben has moral obligations to Vic stemming from BFI.

The search for a demarcation criterion continues, and this is revealed by studying the epistemic case.

4.3. *Congruence and Independence*

I now want to defend the following two claims:

Congruence: For BFI, the type of harm alone dictates the type of obligation.

Independence: Epistemic obligations can arise in the absence of practical obligations, and vice versa.

At play is a rough division of harms into categories such as practical (including monetary), epistemic, emotional, and so on. I make no claim to the depth of these categories in the final accounting, but they suffice for present purposes. The type of harm that has so far dominated discussions about BFI falls into the first category: a person is practically inconvenienced, or left out of pocket. But a person can be unjustly harmed in other ways: emotional abuse can cause harm and be unjust, for example. Congruence says that when the source of a person's moral obligation is that she innocently benefits from injustice it is the type of harm the victim has suffered—rather than, say, the type of benefit gained, or any other factor—which *alone* determines the type of obligation the beneficiary accrues.

We have already seen evidence for Congruence. As noted, it is compatible with the specification of Two Accountants that both applicants are African Americans who differ in degree of similarity to the stereotype. It may therefore be that *both* candidates suffer identity-prejudicial credibility deficit, though Vic more severely than Ben. Ben still acquires moral obligations specifically addressing Vic in his capacity as a knower. The same is true, *mutatis mutandis*, for the case of emotional harm. Together these two cases lend Congruence significant support, since they rule out the type of benefit as a determining factor.

If Congruence is accepted, Independence follows, given the plausible assumption that that it is possible to suffer harm in the epistemic dimension without suffering harm in the practical one—as the Two accountants case demonstrates—and *vice versa*—as the cases in the BFI literature to date evidence. Indeed, a generalised version of Independence will be plausible, unless we find convincing cases where harm in one dimension always goes along with harm in another.

4.4. A General Principle

Finally I want to tentatively suggest the following principle:

Benefitting from Injustice – General Principle: When a person innocently benefits from injustice she thereby acquires a moral obligation to try to make the victim whole again, in whatever dimensions she was harmed, up until a ‘change of positions’.

The General Principle is inherently plausible. It is further strengthened by the fact that it explains and systematises many verdicts about BFI in the literature. For example, in Butt’s island case, C was harmed in the material dimension: she was left without yield because the water source was unjustly diverted. Having benefitted from this injustice, B acquired an obligation to try to make C whole again in that dimension, which entails giving up the 200 extra kilos the injustice delivered her: our verdict exactly.

The Core Claim needs no further support than it receives in §2. Still, the General Principle offers a deeper explanation the it, and of our verdicts in Two Accountants. To try to make something happen is just to act on one’s reasonably held beliefs about what would most efficiently advance that goal, among feasible and permissible actions. To try to make a person whole again in a dimension consists in, and so entails, acting on one’s reasonable beliefs about which actions would most efficiently advance that goal. In Two Accountants, Vic was harmed in the epistemic dimension. It is reasonable for Ben to believe that performing actions which specifically honours Vic in his capacity as a knower, as well as resisting further epistemically disrespectful behaviour, and calling out the original injustice, together will be most efficient at making him whole again. Such actions are feasible, and permissible.³² Thus the Core Claim, (i), and (ii) are all well explained by the General Principle since they follow from it, given plausible empirical assumptions.

The principle also explains the more detailed matter of why there is no particular action among those that would honour Vic which Ben ought to do, and conversely why there’s no leeway in the watercooler situation, or with respect to calling out the committee. The former is because Ben reasonably believes

³² At least insofar as the agent can carry them out without being patronising, overbearing, or in other ways causing harm. I assume that this is possible, and feasible, acknowledging this to be a substantial assumption. [Removed for blind review 9.]

that carrying out any one (or a few) of these actions will have the same effect; the latter is because Ben reasonably believes that if he allows further epistemic dishonour to be inflicted on Vic, or fails to call out the committee, all his other efforts will be in vain.

What of the mentioned 'change of positions'? The obligations BFI generates are widely acknowledged to be limited. The literature has so far largely focused on monetary harm and benefit. Accordingly, the obligation has usually been thought to be limited to relinquishing the material benefit received, or its corresponding market value. Here again we can see the utility of considering epistemic injustice, since this allows us to realise that this cannot be the general principle at play. What Ben needs isn't more money.

What could the principle be? Some have noted that when the benefit is material but not simply money, agents may value it substantially less than its market value (Butt 2007). Moreover, some benefits cannot simply be sold (Cahn 2002). It seems unfair to require an agent to pay more for such a benefit than they would have been willing to pay, even if that sum is substantially lower than the benefit's market value.

Such examples, along with the instances of epistemic injustice we have discussed, indicate that we need a more flexible and general principle. 'Change of positions' fits the bill. It says that obligations generated by BFI are limited at whatever point the beneficiary would rationally have preferred not to benefit to begin with.

This limiting clause is made plausible by its explanatory power. First, it explains the standard cases, in which a person is obliged to pay back the same amount, but not more, than the money they received unjustly: in most circumstances this is the point at which a person would rationally prefer not to have benefited to begin with. Second, it also explains more rarely acknowledged cases in which a person is only obliged to pay back a smaller amount. If Petra innocently receives a thousand dollars at a time when she is affluent but then suffers significant (undeserved and unrelated) financial misfortune, she may fairly reject paying back even half that amount later, on the grounds that money has much higher value for her now. Again this is because it would be rational for her to prefer not to have benefited to begin with.

Finally, and from our perspective most interestingly, the principle works equally well in cases where the benefit and the harm are in different dimensions. In *Two Accountants* there will be a point at which Ben would rationally have preferred not to benefit: the point at which the cost of discharging his obliga-

tions to Vic outweighs the benefit received. There, but not before, is the point at which Ben's moral obligations end. The change of position clause gives an account of the limits of moral obligations arising from BFI that is both principled and generally applicable.³³

*

We have learned various lessons about BFI as a source of moral obligation. First, beneficiaries' obligations are not in general secondary to or derivative of perpetrators' obligations, and the simplest theory says that they never are. Second, the search must continue for a criterion of demarcation, since epistemic injustice provides counterexamples against major contenders. Third, where BFI is the source, the type of harm alone dictates the type of obligation that arises, and fourth, epistemic obligations can arise in the absence of practical obligations, and vice versa. Finally, reflection on the epistemic case leads us to tentatively consider a General Principle for moral obligations stemming from BFI, which limits those obligations in a plausible and generally applicable way.

These insights stem from reflection of the epistemic case. Together with the discussion in §2 they lend significant credence to the more general claim this paper seeks to establish, namely that considering epistemic injustice allows us learn useful theoretical lessons about BFI as a source of moral obligation.

5. Significance

It might be supposed that in cases of epistemic injustice, the norm is that all parties lose, that this gives rise to a tendency to self-correction, and that, consequently, cases of benefit are contrived, rare, and of marginal importance.³⁴ After all, not only does S suffer epistemically; *H does too*: she loses out on a source of knowledge she otherwise would have had. So all parties have an incentive to avoid such situations, and the phenomenon will die a natural death, or so the thought goes.

This is a mistake.

³³ We bracket throughout a person's preference not to benefit from injustice. The limiting clause is generally applicable so long as rationality itself can get a grip on the situation. I have not discussed the reference to what an agent *rationally* would have preferred. Some such proviso is necessary in order that to avoid obligations being limited by bizarre but sincerely held preferences. Spelling this out goes beyond the limits of this paper.

³⁴ [Removed for blind review 10.]

Some take the view that in liberal democracies justice is the norm, and injustice the exception. However, the opposite is more true to the facts: “injustice is ... a normal social baseline, while active cries of resentment and demands for rectification are the precious exception (Fricker 2007, 39).³⁵ Considering specifically the case of epistemic injustice, though many of us would never countenance the conscious operation of identity-prejudice, Fricker rightly puts weight on the point that prejudice can do its dirty work ‘by stealth’, without our knowing, and even contrary to our consciously held beliefs (p. 39).³⁶ There is, as noted, significant evidence of the widespread existence and influence of implicit stereotypes across society.³⁷ Given that we often are in situations of testimonial exchange in which we do not have specific information about the speaker, and given the prevalence of stereotypes, many of which are negative, the most plausible conclusion is, as Fricker also argues, that *epistemic injustice is everywhere*.

This does not yet show that *benefiting* from epistemic injustice is widespread, but this too is plausible. Linda Martín Alcoff has argued that, contrary to the picture drawn by Fricker, *strategic speech*, speech the primary purpose of which is not to transmit information or impart knowledge, but to further some agenda—to sell a product, increase the likelihood of (re-)election, further the interests of an institution, and so on—in fact dominates (Alcoff 2010, 133).³⁸ Speech often forms part of a competition: for status, resources, and so on. When that is the case and someone is not believed, *they* (are more likely to) *lose*. And *then someone else wins*. In the proportion of such cases in which TI is involved, those others benefit from epistemic injustice. Given the prevalence of prejudice and implicit bias, this proportion is probably large.

³⁵ Thus Charles Mills, in his discussion of Goldman’s *Knowledge in a Social World* (1999), writes: “The picture of “society” [Goldman] is working with is one that—with perhaps a few unfortunate exceptions—is inclusive and harmonious. Thus his account offers the equivalent in social epistemology of the mainstream theorizing in political science that frames American sexism and racism as “anomalies”: U.S. political culture is conceptualized as *essentially* egalitarian and inclusive, with the long actual history of systemic gender and racial subordination being relegated to the status of a minor “deviation” from the norm Obviously such a starting point crucially handicaps any realistic social epistemology, since in effect it turns things upside down. Sexism and racism, patriarchy and white supremacy, have not been the *exception* but the *norm*” (Mills 2007, 17).

³⁶ See again n. 8, and §1.2 above.

³⁷ Also from within the target groups themselves; bias among black people against black people, or among women against women, for example (Nosek et al. 2007).

³⁸ Alcoff argues that Fricker overestimates the importance of identity prejudice for credibility deflation, and that in many instances credibility judgements are *duly* deflated because of healthy scepticism to the speaker’s agenda. We can understand this as reasons, perhaps quite general reasons, to believe that the speaker is unlikely to be sincere to the hearer in conditions such as those prevailing. But even if lowering credibility judgements is often legitimate *it does not follow* that credibility judgements are not also often *unduly* deflated by the operation of identity prejudices. Our reasons for thinking that testimonial injustice is widespread are in fact wholly unaffected by this line of reasoning. Fricker’s reply is in her (2010).

We have seen that innocently benefiting from epistemic injustice can generate moral obligations to perform actions that specifically address a person in their capacity as a knower, and that studying the case of epistemic injustice is useful when the aim is to correctly theorise BFI as a source of moral obligation. Those concerned with injustice have an additional reason to pay attention, since benefiting from epistemic injustice is not a marginal and esoteric phenomenon, but a pervasive, central and important one.

References

- Alcoff, Linda Martín. 2010. "Epistemic Identities." *Episteme* 7 (2): 128–37.
<https://doi.org/10.3366/E1742360010000869>.
- Anwander, Norbert. 2005. "Contributing and Benefiting: Two Grounds for Duties to the Victims of Injustice." *Ethics & International Affairs* 19 (01): 39–45.
<https://doi.org/10.1111/j.1747-7093.2005.tb00488.x>.
- Barry, Christian, and Robert Kirby. 2015. "Scepticism about Beneficiary Pays: A Critique." *Journal of Applied Philosophy*. <https://doi.org/10.1111/japp.12160>.
- Barry, Christian, and David Wiens. 2014. "Benefiting from Wrongdoing and Sustaining Wrongful Harm." *Journal of Moral Philosophy*, 1–23. <https://doi.org/10.1163/17455243-4681052>.
- Butt, Daniel. 2007. "On Benefiting from Injustice." *Canadian Journal of Philosophy* 37 (1): 129–52.
- . 2009. *Recifying International Injustice*. Oxford University Press.
- Cahn, Steven M., ed. 2002. "Preferential Hiring and Compensation." In *The Affirmative Action Debate*, 68–78. Routledge.
- Coady, David. 2017. "Epistemic Injustice as Distributive Injustice." In *The Routledge Handbook of Epistemic Injustice*, edited by James Kidd, José Medina, and Gaile Jr. Pohlhaus, 61–68. Taylor and Francis.
- Cooley, E., B. K. Payne, and K. J. Phillips. 2013. "Implicit Bias and the Illusion of Conscious Ill Will." *Social Psychological and Personality Science* 5 (4): 500–507.
<https://doi.org/10.1177/1948550613506123>.
- Dovidio, John F., and Samuel L. Gaertner. 2000. "Aversive Racism and Selection Decisions: 1989 and 1999." *Psychological Science* 11 (4): 315–19. <https://doi.org/10.1002/da.20054>.
- Dunham, Jeremy, and Holly Lawford-Smith. 2017. "Offsetting Race Privilege." *Journal of Ethics and Social Philosophy* 11 (2): 1–22.
- Fabre, Cecile. 2007. *Justice in a Changing World*. Polity.
- Fricke, Miranda. 2007. *Epistemic Injustice*. Oxford University Press.

- . 2010. “Replies to Alcoff, Goldberg, and Hookway on Epistemic Injustice.” *Episteme* 7 (02): 164–78. <https://doi.org/10.3366/E1742360010000894>.
- Gawronski, Bertram, William A. Cunningham, Etienne P. LeBel, and Roland Deutsch. 2010. “Attentional Influences on Affective Priming: Does Categorisation Influence Spontaneous Evaluations of Multiply Categorisable Objects?” *Cognition & Emotion* 24: 1008–25. <https://doi.org/10.1080/02699930903112712>.
- Goldberg, Sanford. 2010. “Comments on Miranda Fricker’s ‘Epistemic Injustice.’” *Episteme* 7 (2): 138–50. <https://doi.org/10.3366/E1742360010000870>.
- Goldberg, Sanford C. n.d. “Should Have Known.” *Synthese*. <https://doi.org/10.1007/s11229-015-0662-z>.
- Goldman, Alvin I. 1999. *Knowledge in a Social World*. Oxford University Press. <https://doi.org/10.1080/02691720010008590>.
- Goodin, Robert E., and Christian Barry. 2014. “Benefiting from the Wrongdoing of Others.” *Journal of Applied Philosophy* 31 (4): 363–76. <https://doi.org/10.1111/japp.12077>.
- Greenwald, Anthony G., T. Andrew Poehlman, Eric Luis Uhlmann, and Mahzarin R. Banaji. 2009. “Understanding and Using the Implicit Association Test III. Meta-Analysis of Predictive Validity.” *Journal of Personality and Social Psychology* 97 (1): 17–41. <https://doi.org/10.1037/a0015575>.
- Greenwald, Anthony G., Debbie E McGhee, and Jordan L. K. Schwartz. 1998. “Measuring Individual Differences in Implicit Cognition: The Implicit Association Test.” *Journal of Personality and Social Psychology* 74 (6): 1464–80.
- Hahn, Adam, Charles M Judd, Holen K Hirsh, and Irene V Blair. 2014. “Awareness of Implicit Attitudes.” *Journal of Experimental Psychology: General* 143 (3): 1369–92. <https://doi.org/10.1037/a0035028>.
- Haydar, Bashshar. 2003. “The Moral Relevance of Cost.” *Philosophical Studies* 112 (2): 127–34. <https://doi.org/10.1007/S11098-008-9324-Z>.
- Haydar, Bashshar, and Gerhard Øverland. 2014. “The Normative Implications of Benefiting from Injustice.” *Journal of Applied Philosophy* 31 (4): 349–62. <https://doi.org/10.1111/japp.12078>.

- Holroyd, Jules. 2012. "Responsibility for Implicit Bias." *Journal of Social Philosophy* 43 (3): 274–306. <https://doi.org/10.1111/j.1467-9833.2012.01565.x>.
- Hookway, Christopher. 2010. "Some Varieties of Epistemic Injustice." *Episteme* 7 (02): 151–78. <https://doi.org/10.3366/E1742360010000882>.
- Huseby, R. 2013. "Should the Beneficiaries Pay?" *Politics, Philosophy & Economics*, 1–17. <https://doi.org/10.1177/1470594X13506366>.
- Jost, John T., Laurie A. Rudman, Irene V. Blair, Dana R. Carney, Nilanjana Dasgupta, Jack Glaser, and Curtis D. Hardin. 2009. "The Existence of Implicit Bias Is Beyond Reasonable Doubt: A Refutation of Ideological and Methodological Objections and Executive Summary of Ten Studies That No Manager Should Ignore." *Research in Organizational Behavior* 29: 39–69. <https://doi.org/10.1016/j.riob.2009.10.001>.
- Lackey, Jennifer. n.d. "Norms of Credibility." *American Philosophical Quarterly*.
- Langton, Rae. 2010. "[Review of the Book Epistemic Injustice: Power and the Ethics of Knowing, by Miranda Fricker]." *Hypatia* 25: 459–64.
- Lawford-Smith, Holly. 2016. "Offsetting Class Privilege." *Journal of Practical Ethics* 4 (1): 23–51.
- Linville, P W, G W Fischer, and P Salovey. 1989. "Perceived Distributions of the Characteristics of In-Group and out-Group Members: Empirical Evidence and a Computer Simulation." *Journal of Personality and Social Psychology* 57 (2): 165–88. <https://doi.org/10.1037/0022-3514.57.2.165>.
- Maddox, K. B., and S. a. Gray. 2002. "Cognitive Representations of Black Americans: Reexploring the Role of Skin Tone." *Personality and Social Psychology Bulletin* 28 (2): 250–59. <https://doi.org/10.1177/0146167202282010>.
- Maddox, Keith B. 2004. "Perspectives on Racial Phenotypicality Bias." *Personality and Social Psychology Review* 8 (4): 383–401.
- Madva, Alex. 2017. "What Can We Learn from the Implicit Association Test? A Brains Blog Roundtable." *The Brains Blog*. 2017.
- Maitra, Ishani. 2010. "The Nature of Epistemic Injustice." *Philosophical Books* 51 (4): 195–211.
- Miller, David. 2001. "Distributing Responsibilities." *Journal of Political Philosophy* 9 (4): 453–71.
- Mills, Charles W. 2007. "White Ignorance." In *Race and Epistemologies of Ignorance*, 11–38.

- <https://doi.org/10.1080/02691721003749927>.
- Narveson, Jan. 2003. "We Don't Owe Them a Thing! A Tough-Minded but Soft-Hearted View of Aid to the Faraway Needy." *The Monist* 86 (3): 419-33.
- Nosek, Brian a., Frederick L. Smyth, Jeffrey J. Hansen, Thierry Devos, Nicole M. Lindner, Kate a. Ranganath, Colin Tucker Smith, et al. 2007. "Pervasiveness and Correlates of Implicit Attitudes and Stereotypes." *European Review of Social Psychology* 18 (1): 36-88.
<https://doi.org/10.1080/10463280701489053>.
- Oswald, Frederick L, Gregory Mitchell, Hart Blanton, James Jaccard, and Philip E Tetlock. 2013. "Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies." *Journal of Personality and Social Psychology* 105 (2): 171-92.
<https://doi.org/10.1037/a0032734>.
- Pasternak, Avia. 2014. "Voluntary Benefits from Wrongdoing." *Journal of Applied Philosophy* 31 (4): 377-91. <https://doi.org/10.1111/japp.12072>.
- Payne, B K. 2001. "Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon." *Journal of Personality and Social Psychology* 81 (2): 181-92.
<https://doi.org/10.1037/0022-3514.81.2.181>.
- Payne, Keith, and Kristjen Lundberg. 2014. "The Affect Misattribution Procedure: Ten Years of Evidence on Reliability, Validity, and Mechanisms." *Social and Personality Psychology Compass* 8 (12): 672-86. <https://doi.org/10.1111/spc3.12148>.
- Singer, Peter. 1972. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1 (3): 229-43.
- Smith, Colin Tucker, and Brian a. Nosek. 2010. "Implicit Association Test." In *Corsini's Encyclopedia of Psychology*, 4th ed., 1-4. Wiley. <https://doi.org/10.1016/B978-0-08-097086-8.25069-1>.
- Steinpreis, Rhea E., Katie A. Anders, and Dawn Ritzke. 1999. "The Impact of Gender on the Review of the Curricula Vitae of Job Applicants and Tenure Candidates : A National Empirical Study." *Sex Roles* 41 (7/8): 509-28.
- Stroud, Sarah. 2006. "Epistemic Partiality in Friendship." *Ethics* 116 (3): 498-524.
<https://doi.org/10.1086/500337>.

- Thomson, Judith Jarvis. 1973. "Preferential Hiring." *Philosophy & Public Affairs* 2 (4): 364-84.
- Woodward, James. 1986. "The Non-Identity Problem." *Ethics* 96 (4): 804-31.
- Worsnip, Alex. 2014. "Disagreement about Disagreement? What Disagreement about Disagreement?" *Philosophers' Imprint* 14 (18): 1-20.