

Benefiting from Epistemic Injustice

ABSTRACT

This paper argues that benefiting from epistemic injustice in some cases generates stringent moral obligations with distinctively epistemic content. It further argues that considering the case of epistemic injustice allows us to better appreciate how benefiting from injustice, as a source of moral obligation, should best be theorised more generally.

Two recent developments in political theory are the following. On the one hand, Miranda Fricker (2007) has argued that people suffer a distinctively epistemic form of injustice when they are harmed in their capacity as a knower. On the other hand, a range of authors have joined Daniel Butt (2007) in arguing that benefiting from injustice ('BFI') can be a source of moral obligation. In this paper I put these two ideas together, and ask which obligations arise when a person benefits from *epistemic* injustice ('BFEI').

The paper's primary aim is to establish what I shall call its Core Claim, namely that in some such cases, the beneficiary of epistemic injustice acquires stringent moral obligations with a specifically epistemic content. The secondary aim is to establish that for BFI, the type of harm alone dictates the type of obligation incurred, and that epistemic obligations can arise in the absence of practical obligations, and vice versa. Finally, the paper's more general aim is to demonstrate by example the theoretical fecundity of coupling these two debates. The literature on BFI has been dominated by consideration of cases where the harm suffered is practical (often monetary). As we shall see, BFEI involves harm of a different sort, and therefore proves especially useful to our efforts to correctly theorise BFI as a source of moral obligation.

A caveat. My focus will be on the actions a beneficiary of injustice is morally obliged to perform *vis-à-vis its victim*. I do not mean to suggest that these are the only moral obligations accrued in cases of BFEI. As has often been pointed out in critique of Fricker, epistemic injustice has structural causes and likely requires structural reform to properly remedy. In benefiting from injustice of any kind, one plausibly acquires moral obligations to address the structural causes of those injustices. So, too, in the case of epistemic injustice. These are important and weighty obligations, but they will not be

addressed here. The obligations we have directly to one another are important too, and there will be more than enough to say about them.

1. Epistemic Injustice

This section discusses some central features of Miranda Fricker's theory of Testimonial Injustice ('TI'), the focus of her recent book *Epistemic Injustice* (2007).¹ Let's begin with a simple example.

Suppose that Sally tells Holly that Arsenal won on the weekend mainly due to Iwobi's outstanding performance. Holly already thought that the win was most likely ascribable to Iwobi; she held a credence of .6 in this proposition. As a result of hearing Sally's testimony, she increases her credence to .7. Simplifying somewhat, Sally suffers an epistemic injustice in this situation according to Fricker if Holly *would have* increased her credence in the testified proposition more but for a prejudice she holds against a group of which she takes Sally to be a member. Let's suppose Holly holds the prejudiced view that women cannot properly analyse football matches, and that she would have adjusted her credence to .85, if the testimony had originated from any of her male friends equally interested in football. Then Sally suffered an epistemic injustice. Schematically we can say that a speaker S suffers TI just in case a hearer H increases her confidence in a proposition to which H testifies *less* than H otherwise would have because H is prejudiced against the social group to which H takes S to belong.

This is fine as a first pass, but to properly evaluate Fricker's position we need to understand it in more detail. Let's begin by considering the situation from the point of view of the hearer. As in our example, let p be a proposition H has already entertained, and which she believes to some degree.² S's saying that p constitutes evidence for p . This means that H is now faced with the *rational challenge* of deciding how much more strongly she should believe that p , given this new evidence in its favour. If she already had high credence in p , new evidence will tend to move her credence very little. If her prior credence was low, this new evidence might move her credence quite a lot.³

¹ In the book, Fricker discusses two forms of epistemic injustice: TI 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 this book, unless context makes clear otherwise.

² On quite standard theories of credences, agents are assumed to hold credences for all propositions.

³ Though perhaps not if the prior credence is *very* low. See (Jones 2002), and discussion below.

H's immediate predicament is to update her credence, such as it actually is. However, she might easily have had a different prior⁴ credence in p . What we really need, therefore, is a function which accepts as inputs *any* prior credence in p , as well as the evidence that S says that p here and now, and yields as output a posterior credence in p . Such a function would constitute a more general answer to the question of how H should adjust her credence in p , given that S has just said that p . Hence we might aptly describe the rational challenge H faces as that of *selecting the correct updating function* for her credence in p .⁵

Fricker notes two types of considerations which have particular relevance for this rational challenge. The first is H's assessment of S's *competence* with respect to p . The second is H's assessment of S's *sincerity* with respect to p (to hearers like H, in situations like these). In some cases, a hearer has an abundance of information that can help her to rationally assess these questions. We know some people's areas of competence well, and sometimes we know things about them that enable us to understand what incentives they have to be sincere—or otherwise.

But very often this is not the case. In many testimonial exchanges we know very little about the speaker. We may have just met them for the first time, and have no prior knowledge about them. In that case we know nothing about S's competence beyond what we know about the general level of competence regarding such propositions in the population at large. We also know nothing about S's incentives to be truthful or to lie beyond our knowledge of incentives everyone has. So we have very little to go on, and yet update our credence we must, since not changing one's credence in the face of new evidence is still to have decided on an updating function.

Fricker argues that an unobjectionable response to this predicament, both from the point of view of rationality, and from the point of view of morality, is for the hearer to rely on *stereotypes*: widely held associations between social groups and one or more traits (p. 30). For one, given that we often lack specific information about a person, doing so is *inevitable*. Moreover, many generalisations are *accurate*, such as the generalisation that a family doctor is competent and trustworthy about common ailments and diagnoses. So Fricker argues that the use of stereotypes to make *credibility judgements*—judgements about speakers' competence and sincerity—is in itself unobjectionable.

But not every instance of it is. Some stereotypes constitute *prejudices*, and when prejudices operate, the result can be TI. What are prejudices? To begin with, all prejudices are *inaccurate* stereotypes, but not all inaccurate stereotypes are prejudices. This is (among other reasons) because inaccurate

⁴ Prior to the new evidence coming in, and to updating therefore being required.

⁵ The reconstruction of this challenge in this and the previous paragraph is my own, not Fricker's.

stereotypes can result from honest mistakes, for example due to misleading data, such as non-representative samples.⁶ So how do we tell an inaccurate stereotype from an honest mistake?

Fricker accepts Nomy Arpaly's view that a good test is what a person does when he or she is exposed to corrective data. Arpaly gives the example of Solomon, a boy from a small, isolated community in which all skilled abstract thinkers are male, all books about abstract matters have male authors, etc. (Arpaly 2000, p.103). Solomon believes that women either aren't skilled abstract thinkers, or perhaps are just not inclined towards such pursuits. This might be an honest mistake, given the skewed sample to which he has been exposed. However, Solomon later attends university in the city, and is promptly exposed to numerous examples of skilful abstract thinkers that are female. Arpaly argues that if Solomon was honestly mistaken before, he will now change his view, but if he persists in his belief, it is revealed to not have been an honest mistake to begin with.

Retaining belief in the face of (sufficient) counterevidence renders H epistemically blameworthy. However, holding an inaccurate stereotype despite contrary evidence is still not sufficient to count as harbouring a prejudice, on Fricker's view. On her account, for something to count as a prejudice, the explanation for *why* the stereotype resists modification must be a certain type of *affective investment* (p. 35). This is, on Fricker's view, where *moral* blame enters the picture, since some such affective investments are themselves ethically noxious. In the example of Solomon, as well as in the example of Holly and Sally above, this ethically noxious motivation might be contempt for women, for example (p. 34). So, according to Fricker, a prejudice is:

Prejudice: an inaccurate stereotype—that is, an inaccurate generalisation between a social group and one or more traits—which persists despite counterevidence because of an ethically noxious affective investment

An example is the false generalisation that women are not skilled abstract thinkers, if it persists despite counter-evidence because of contempt against women.

A testimonial injustice occurs when H's credibility assessment of S is unduly deflated as a result of the operation of a prejudice, so conceived: H assesses S's competence, sincerity, or both, less favourably than the evidence indicates (hence *unduly* deflated), because of the operation of a prejudice. In doing so, H does S an injustice, and H is "*wronged specifically in her capacity as a knower*", according to Fricker (p. 20).

⁶ My interpretation of Fricker here differs from that of Ishani Maitra (2010, p.197), who thinks a stereotype being accurate is compatible with it counting as a prejudice on Fricker's account.

TI: A hearer's credibility assessment of a speaker is unduly deflated because of the operation of a prejudice

TI can occur in many different types of cases, but some are more important than others. That's because some ways of placing people in categories operate in restricted settings, while others have wide application. Imagine that football players generally regard strikers as being full of themselves. This might affect strikers on the field and in related social settings, but it won't reach much further. Stereotypes that take point of departure in such categories are thus of limited effect. By contrast, other social categories are operative in a wide range of settings: examples include (perceived) race or ethnicity, gender, and sexual orientation. These are what Fricker calls *tracker* prejudices; prejudices which are likely to track a person through a variety of different kinds of interactions: "economic, educational, professional, sexual, legal, political, religious, and so on" (p. 27).

The social categories which track people across settings in this way are usually (always? See p. 27) constitutive of a person's *social identity*. When a person suffers TI linked to a one of the categories that constitute identity, this usually goes along with a wide range of other injustices. Testimonial injustices which are connected with other injustices are *systematic* (p. 27). Identity prejudices are, moreover, *persistent*, because the prejudices that underlie them change slowly if at all, so the person is typically subjected to such injustices their entire life. Hence, while testimonial injustice can take point of departure from other social categories too, the central case, given our interest in injustice, is *credibility deficit due to identity prejudice*.

2. Updating the Account

I now wish to discuss some ways in which I believe this theory needs amendment. This is no attempt at comprehensive review of the theory. There are aspects of it that to my mind are implausible or ill motivated, but which for our purposes can be bracketed.⁷ There are also many critical engagements

⁷ An example is Fricker's insistence that prejudicial stereotypes operate "without doxastic mediation" (p. 36). The argument for this seems to be that the widely acknowledged point that the content of stereotypes may be inconsistent with a person's beliefs. But this is easily accounted for on 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. So I can see no motivation to move to an account of a stereotype as "an *image* which expresses an association between a social group and one or more attributes" (p. 37). But nothing in this article hinges on this point.

with the book to which I think Fricker has convincingly replied, or which (again) do not seem relevant for our purposes. These will also be set aside here.

What hearers owe to speakers

The first modification I think is necessary and relevant is due to Ishani Maitra (2010). She argues that many cases where credibility judgements are unduly deflated due to negative identity-prejudices nevertheless fail to constitute testimonial injustices.

To show this, Maitra discusses Zara, a person who disregards a commentary article from a 'tea-partier' simply on the basis of prejudice against that group of people. Zara *never* reads articles from 'tea-partiers', so she has blocked herself from counter-evidence against her stereotype. But evidence is readily available to her: many tea-partiers are knowledgeable about the issues on which they comment. Had she correctly assessed the evidence, she would not have dismissed them all out of hand. Nevertheless, Maitra argues, Zara commits no injustice to the writer in dismissing him, because in simply browsing the paper she "doesn't owe it to him to avoid a credibility deficit" (Maitra 2010, p.199). We owe some *but not to all* speakers to make correct credibility judgements.

One of Maitra's arguments here parallels Fricker case for the acceptability of using stereotypes in the first place. Given the amount of information we are subjected to, and the scarcity of time and other resources available, 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). Many sources of testimony engage with us without invitation, indeed often against our preferences—consider billboards, TV and radio advertising, people speaking loudly on public transport, and so on. So it is implausible that we owe these people the resources it would take us to make accurate credibility-judgements. We have "no *general* obligation" to make correct credibility judgements, Maitra argues, so accusations of TI must be backed up by an argument that in the case at hand, such an obligation exists (or by a general theory to that effect).

Proper regard for the evidence

The second amendment to the account has the same origin. Fricker's official formulation of a prejudice employs the notion of (culpable) 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).

The key point is that one can fail to have proper regard to the evidence in other ways than by resisting counter-evidence. For example, one can generalise to a wide group of people on the basis of an absurdly small sample (arguably, as she notes, at issue in the case of Solomon), or one can fail to

seek out easily available evidence about the reliability or otherwise of one's stereotypes. Maitra seems clearly correct in claiming that these failures to give proper regard to the evidence make the generalisations prejudicial. And this is so even if, *had* one received that counter-evidence, one *would* have changed one's mind (so that the generalisations are not held with resistance to counter-evidence, and so do not qualify as prejudices on the official formulation; Maitra 2010, pp. 205–7).

Implicit Bias

The final amendment I want to suggest to Fricker's account is aimed at allowing it to accept the support it intuitively receives from recent research in psychology. Over the past few decades there has been a burgeoning of work aimed at measuring *implicit bias* (IB) and its effects. Simplifying greatly, we can think of IB as an inaccurate negative stereotype a person is unaware of harbouring.⁸ For example, many readers 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 §8 below.) Empirical research has shown that IB is most likely widespread.⁹ It is natural to take the latter fact to strongly support the former

⁸ With somewhat more precision, *S* can be said to hold IB 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 IB, but takes a stance on its explanatory power.

⁹ Some will regard this statement with scepticism or incredulity, aware as they are of recent research casting doubt on the predictive power of the Implicit Association Test (IAT), a prominent and recently much-discussed research paradigm on IB (for example, Oswald et al. 2013). However, (1), as Alex Madva has recently 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 IB on the basis of challenges facing *one particular research paradigm* which aims to study it, is obviously unwarranted. IB can and 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). Consider, to take just one example, the seminal study by Steinpreis, Anders, and Ritzke (1999), who sent identical copies of a CV—half marked with a (fictitious) male name, half with a 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 hire ‘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 clearly indicate both the existence and causal efficacy of implicit bias—unless one is prepared to adopt the utterly implausible hypothesis that a sufficient

contention.¹⁰ Unfortunately, Fricker's account doesn't admit of this support as it stands.¹¹ That's because TI arises when credibility is depressed due to *prejudice*, and many cases of implicit bias will fail to count as prejudices, on her view. Recall that Fricker takes a stereotype to constitute a prejudice only if its resistance to counterevidence is explained by a *morally noxious motivation*, such as contempt for women. For many examples of implicit bias, this explanation is implausible. Consider implicit preference for young people over old in domains where age has no importance (or where age, to the contrary, is an indicator of valuable experience), for example. This stereotype's

fraction of the participants in fact were *explicitly* biased against women (as well as coming up with 'debunking' explanations for the numerous other studies discussed in this paper).

Another important paradigm for the study of Implicit Bias is the Affect Misattribution Procedure (AMP), in which 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 relevant dimension; Chinese pictographs are ambiguously 'pleasant / unpleasant' for those who cannot read them, for example, and any residual inherent 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. 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, crucially for present purposes, indeed *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 & Lundberg 2014). This paradigm has e.g. demonstrated significant ($p < 0.001$) implicit bias against homosexual men / couples (Cooley et al. 2013), blacks, and the elderly ($p < .05$ in both cases) (Gawronski et al. 2010), and this paradigm has been found to predict behaviour, better, for example, than IATs do (for discussion, see Payne & Lundberg 2014, p.674).

For an introduction to IATs, see for example Smith and Nosek (2010) or (Greenwald et al. 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 IAT'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 & Gaertner 2000). On the other hand, for an argument that the states tested by IATs are not after all ones of which the agent is unaware—and thus not *implicit* biases—see (Hahn et al. 2014).

¹⁰ Why focus on *implicit* bias? Wouldn't the contention be equally well supported if bias of whatever kind is widespread? Yes and no. If we could show that explicit tracker prejudice is widespread, we would get the support. But most do not believe that it is: they do not take their friends and relatives to be *explicitly* biased against people on the basis of their skin colour, sexual orientation, etc. The preponderance of implicit bias therefore lends Fricker's contention much stronger support.

¹¹ Thanks to NN4 for this point.

failure to be sensitive to evidence seems ill explained by a morally noxious motivation. A more plausible explanation is our culture's constant reinforcement of the ideal of youth, for example through advertising and other pop-cultural expressions. Yet, in a significant proportion of the cases where S receives a deflated credibility assessment due to the operation of this stereotype, S surely suffers an epistemic injustice. So we need an extension of the account.

Accordingly, I suggest that any case in which a speaker suffers a credibility deficit on account of implicit bias, and where the hearer is morally blameworthy¹² either for harbouring the bias, or for allowing it to have practical effects, or for both, shall count as a case of TI.

This proposal has clear benefits. On the one hand, it allows us to respect the spirit of Fricker's account, in that moral noxiousness remains a precondition of TI. We will *rule out* a number of cases of implicit bias causing credibility deficit, as I believe is proper. On the other hand, we can also *rule in* a number of such cases, and many more than does Fricker's unamended account. At least to the extent that the prevalence of implicit biases is widely known, and insofar as it is possible, moral agents acquire moral obligations to do what they can to limit the effects of such biases by their own hand.¹³ Failure to do so will suffice for the moral noxiousness we need. In this way, the stance that TI is widespread can accept the empirical support it intuitively gets from the research on implicit bias, while making sensible and balanced judgements about the various different cases.

Summary

In what follows, therefore, I assume an account of TI like Fricker's, except amended as follows. *First*, there is TI only when H fails to discharge an obligation to make a correct credibility judgement.¹⁴ *Second*, I adopt the more demanding understanding of what 'proper regard for evidence' means—more demanding, in particular, than merely being receptive to counter-evidence should one happen upon it. *Third*, credibility deficit can qualify as TI without the stereotype's inadequate sensitivity to

¹² I leave it open what the conditions for such blameworthiness are; see (Holroyd 2012) for discussion.

¹³ 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 IAT 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, p.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.

¹⁴ Describing the correct account of when this obligation arises is a task I must leave for another occasion.

the evidence being explained by morally noxious *motivation*, provided that the deflated credibility judgement is due to the operation of implicit bias for which the hearer bears moral responsibility.

3. The Harm of Testimonial Injustice

The above amendments do not take issue with Fricker's claim that when a credibility judgement is unduly deflated, S is harmed specifically in her capacity as a knower. Nor will I engage critically with *this* part of Fricker's view in this paper: my aim is to explore the consequences of accepting it 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 harm is extrinsic. Let's take them in order.

What gives human life its special value? On many accounts, an important part of the answer is the human capacity for *reason*. To be a person capable of knowing, and of transmitting information, is on views of this kind an important part of our humanity. To unduly deflate one's credibility assessment of a speaker is therefore, on Fricker's account, to *insult and dishonour* the person in a respect central to her full worth as a human being.¹⁵

It is important to note that this injustice is *intrinsic*: if an identity-prejudicial deflation of a credibility judgement occurs, then on Fricker's view, an injustice has *ipso facto* occurred. This may seem to stretch the concept of injustice.¹⁶ For suppose that H hides her feelings well, and never acts on her deflated assessment. If the assessment has no consequence, where's the injustice?

We need not get too hung up on this question. Again, my purpose is to explore how the theory of epistemic injustice interacts with that of BFI. 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 will continue to speak of injustice, but readers should feel free to substitute with 'wrongdoing' as appropriate.

¹⁵ When H unduly deflates her credibility assessment of S, Fricker claims that she, in an important sense, treats S as *less than fully human*. Further, that S 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.

¹⁶ Thanks to NN1 for discussion.

The secondary epistemic injustice relates to downstream consequences of the deflated credibility assessment. Recall that the central case of TI involves an identity-prejudicial stereotype, and Fricker's point that such injustice tends to be persistent and systematic. As a result of such treatment a person can lose confidence in her own beliefs. Since sufficient confidence is a requirement for belief (on many accounts), the person might end up with a diminished stock of knowledge.

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 of her gender. 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.)

To sum up, the harms from TI come in two categories: intrinsic and extrinsic. The intrinsic harm occurs simply in virtue of the credibility assessment's deflation due to an identity prejudice; it constitutes *in and of itself* an insult and a dishonour to the speaker in a capacity that is central to human value. The extrinsic harms can diminish S's stock of knowledge (or justified belief¹⁷), or degrade or stunt the development of the person as an epistemic agent.

4. Benefiting from Injustice

It's uncontroversial that one's own 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. If I harm you, I acquire a moral obligation to reverse the harm, with a similar proviso. Another category of uncontroversial moral obligations stems from *relationships*: being a parent or guardian gives rise to moral obligations, for example.

Slightly more controversially, one can have moral obligations in virtue of one's abilities. If I am able to save the person suffering cardiac arrest right in front of me at the cost of being late for dinner, most will readily agree that I ought to do so. Since I am analogously able to prevent the suffering of some

¹⁷ Although current orthodoxy in epistemology holds that knowledge is what really matters, other views are available. For example, Kaplan (1985) argues that justified belief, and not knowledge, is the true gold standard. Taking such a view will not impact the present point, however, since epistemic injustice will just as plausibly reduce the agent's stock of justified beliefs. So long as one takes there to be a need for a binary concept of belief, in addition to a concept of credence (Tang 2009 argues that no binary concept is required), it remains plausible that a certain level of confidence is required for belief to obtain, so the lesson transfers directly.

of the very poor at low or negligible cost to myself, it is plausible that I have an obligation to do just that (Singer 1972). The key question about BFI 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) introduces the following thought experiment.¹⁸ On an isolated island, and separated from each other, live farmers A, B, C and D. Their yield depends on two factors: water, which an underground river distributes evenly on the island, and their efforts. 200 kilos suffices for necessities, and five hours of daily work ensures this yield. B, C and D choose this course. By contrast, A works all day, harvests 700 kilos, and enjoys leisure products and abundant food derived from the crops. At some point, D sets out to covertly redirect the river away from B and C, toward her own property, hoping to harvest more while working less. The plan fails; D accidentally redirects the water from herself and C toward B, and subsequently kills herself in despair. (Suicide is rife in this literature.) We're thus left with A, who worked hard and harvested 700 kilos; B, who innocently benefitted from D's actions by harvesting 400 kilos from the standard five hours of work; and C, who also worked five hours a day, but who now has nothing to show for it (Butt 2007, p.132).

C will perish if the situation persists, so *someone* should help her. The only question is *who* (Miller 2001). A framework which sees an agent's actions, relationships, and abilities as the only sources of moral obligation would say that the burden falls to A, in virtue of her greater capacity. But intuition pulls in a different direction. C is the victim of an injustice of which B is the direct beneficiary. It seems that the obligation is on B to relinquish these benefits to save C.

In fact, we need nothing as complicated as this example to motivate the case. If A steals B's car, gives it to C, and then flees leaving no assets behind, C is morally obliged to return the car to B, in virtue of having benefited from the unjust theft. However, Butt's example also supports the stronger conclusion that sometimes benefiting from injustice outranks other sources of moral obligation.

Some theorists are sceptical of the claim that BFI generates moral obligations. However, even the most vocal among them agree there are no conclusive arguments to this effect (Anwander 2005, p.41). We should not, I think, understand this merely as an acknowledgement of the scarcity of proofs of non-existence: it's hard to think of considerations which could undermine the clear verdicts intuition yields in cases such as the ones we've mentioned. The strategy has often been to argue that BFI goes along with something else; there's a hidden element of contributing or upholding the

¹⁸ 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. An important early paper is Thomson's 'Preferential Hiring' (Thomson 1973). Butt's paper clearly did a lot to revitalise the issue in current debate, however.

wrong, for example, and that this is the real source of the obligation (Anwander 2005; Barry & Wiens 2014). Our verdict that there's something wrong in BFI in itself is explained away.

I think this strategy fails. Butt's island case rules out, for example, that it's a person's increased ability to help which is the real source of obligation, for had that been the case we would have said that A, with his very abundant 700 kilos, should be the one to fork out. Luck egalitarianism has also been suggested as an alternative explanation (Huseby 2013), but cannot explain the directed nature of the obligations (Barry & Wiens 2014). And property rights, which might do well in the case of car theft, fail in the island example. Even thinking individually about cases, then, things don't look good for this strategy.

There are also more systematic answers available. First, the reply just considered cannot explain our strong sense that these cases share morally relevant features, because even were it to succeed in a piecemeal fashion, it would thereby become committed to the negation of this claim. But the cases *do* share morally relevant features. Second, BFI as a source of moral obligation is supported by consideration of cases, and our intuitive reactions to them. The explaining away 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 benefitting from the injustice* that they arise.

So I think resistance to BFI as a moral obligation ultimately fails. Moreover, my aim here is to explore how consideration of epistemic injustice sheds light on how BFI should best be theorised. This question is independent of and prior to the question of whether BFI *ultimately is* a source of moral obligation since, to properly evaluate whether BFI is a plausible candidate for a source of moral obligation we must first identify the most plausible theory of it as such. For these reasons, then, I will engage no further with the contention that BFI is *never* a source of moral obligation, but take it for granted that it is.

A more interesting question is what the criterion is for demarcating cases where BFI generates a moral obligation from cases where it does not. I take it to be clear that there are cases of the former sort, but clear, too, that there are cases of the latter. If Sarah and Tom both seek refuge after a terrorist attack in a bar and proceed to strike up a profitable business partnership, they benefit from an injustice, but do not seem to have a special obligation to compensate the victims (Barry & Wiens 2014, p.3).¹⁹

¹⁹ The authors attribute the case to Garrett Cullity.

One suggestion, due to Bashar Haydar and Gerhard Øverland (2014), is that that BFI gives rise to moral obligations only when at least one of certain other conditions are also present. The first is when injustice perverts “a more or less structured and relatively fair competitive procedure for allocating a given benefit or award” (p. 354). The second is when the “action were aimed at, and were motivated by, benefiting” the beneficiary ([sic], p. 356). Finally, the last condition is that there is a “transfer of assets from the perpetrators of the injustice ... to the beneficiaries ...” (p. 359).

In contrast, Barry and Wiens (2014) argue that it is only when BFI *sustains a wrongful harm* that it gives rise to obligations. According to these authors, BFI gives rise to moral obligations (of the type they are interested in—remedial duties) just in case at least one of the two following conditions are satisfied.

1. the beneficiary receives an item or quantum of value to which the victim of the wrongdoing has a claim *and* the victim’s claim remains unresolved;
2. the beneficiary receives a benefit in violation of the victim’s claim(s) on the wrongful practice or institution *and* the victim’s claim(s) remain unresolved (p. 14).

We return to the lessons we can learn about BFI more generally by reflecting on the specific case of BFEI below. The takeaway from this section is that BFI is to be regarded, for now, as a source of moral obligation in at least some cases, and that the criterion demarcating cases that do give rise to such obligations from those that do not is still up for debate.

5. Moral Obligations with Epistemic Content

The primary aim of this section is to argue for the paper’s Core Claim. The secondary aim is to draw out some general lessons about BFI from reflection on the epistemic case.

Core Claim: Innocently benefiting from epistemic injustice (‘BFEI’) 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 expresses a claim about what S ’s obligations are in that space of normativity which she inhabits in virtue of being a *knower*. Obligations of this type are also sometimes denoted with talk of what is *rationally required*, or by reference to *rational criticisability* (omitted). The truth of the sentence is compatible with the claim that *all things considered* S ought to have credence y in

p , where $y \neq x$.²⁰ In the Core Claim, the obligations are not epistemic but *moral*, but they have distinctively epistemic *content*. The nature of the moral ought is of course itself deeply contested, but I will assume that we have a sufficiently shared conception of it for discussion to proceed fruitfully. However, the other part of the claim is worth elaborating.

By saying that the moral obligations have epistemic content, I do not mean to defend the view that there are true claims of the form, ‘morally, S ought to believe that p ’. I think it is reasonable to interpret Fricker as being committed to such claims, and similar claims have been discussed by others (Stroud 2006). But I do not intend this. Instead what I have in mind is that the obligations relate to *actions which address their object specifically in that person’s capacity as a knower*. For example, if I help you to understand a logical truth, I have carried out such an action, and likewise if I direct my children’s attention to epistemically significant parts of a scene they are inspecting—the structure of the ant hill, say. The Core Claim of the paper, then, is that innocently benefiting from epistemic injustice can generate moral obligations to perform actions that specifically address a person in her capacity as a knower.

So much for the content of the main claim. How will it be established? I shall argue for the claim by examining a central case:

Two Accountants: A major company is recruiting one junior and one senior accountant. The two positions differ with respect to pay, and time set aside for professional development. Both will be filled through a single process. After an initial, fair round, only two applicants are left, and they are ranked equal first. Neither applicant is formally qualified for the job, but, as is common, the position description explicitly allows for experience to make up for lack of formal qualifications. On paper, the candidates’ experience is equivalent in relevant respects, but there is uncertainty as to how significant it really is. The candidates’ own testimony in the interviews is crucial to this evaluation. The applicants are visibly different in traits relevant to perceived ethnic identity. As a result of the operation of identity-prejudice, Vic receives a significant credibility deficit, and his account of the depth of his experience (beyond what is documented on paper) is given low credence. He therefore gets the junior position. His competitor Ben’s corresponding account is given higher

²⁰ For example, it may be that S ought epistemically to have credence .2 that she will win the race, but ought prudentially and (therefore) all things considered to have credence .9 in that proposition (Sosa 2007b).

credence, and he gets the senior one. The situation later becomes transparent to both, but there is nothing they can do to reverse the process, start it anew, or to bring about institutional change.²¹ It is also clear that the perpetrators of this injustice will do nothing about it.

In 'Two Accountants', a person suffered an epistemic injustice. As a result, an innocent person benefited. What are the consequences of this for what they owe to each other?

Let's begin with what Ben owes Vic in practical terms. Since he secured the benefit of a better paying position because identity-prejudice tilted the scales in his favour, it seems clear that he must do what he can to counteract the consequences of the injustice. Neither person knows what would have happened had identity-prejudice not come into the mix. But they do know that it did, and that before it did, they both had an equal shot at the senior job. It seems that the just thing to do, given that they cannot force a re-evaluation, would be to share the benefits of the senior position equally. Ben thus has a moral obligation to share the pay gap between the two positions with Vic, so that the two end up with equal remuneration. Moreover, Ben owes it to Vic to try to make up for the latter's absent opportunities for professional development. That may not be possible to achieve completely, since Ben but not Vic is entitled to use work-time to learn and develop professionally. But he may succeed in partially levelling the balance, for example by tutoring Vic in their joint spare time.

These practical obligations are of course *pro tanto*; and other considerations may outweigh them. Indeed, the just thing would be to share the benefits *and burdens* of both positions equally, but that will likely prove impossible in practice. For example, the senior position may require longer hours of work than the junior one, making and being accountable for more difficult decisions, being 'on call' even when off duty, or in other ways be more arduous. This would mitigate Ben's responsibility to share the pay. Further, completely compensating Vic for this loss in professional development may turn out to be so onerous given Ben's other obligations that he would prefer the junior position to doing this, and this would also mitigate our judgement. This is what's been called the *change of*

²¹ This assumption is merely simplifying, and nothing in the argument relies on it. That is because having the type of obligations which are my focus here obviously does not exclude having other obligations, for example to effect structural change. It is not an implausible assumption, however. It is common for the stratum of people who makes decisions to be so radically separated from the rank and file of employees that there is virtually no influence from the latter group to the former—think of high powered law firms, for example.

position defence (Goodin & Barry 2014, p.369), and we will return to it later. For the moment let us note that, at least on the face of things, these seem to be Ben's practical obligations.²²

So far, the account is quite standard for cases of BFI. However, in addition to the practical harms, Vic has also suffered distinctively epistemic harms. On the account we have adopted, Vic suffered both a primary and a secondary epistemic harm. The primary harm obtains simply in virtue of the credibility deficit he suffered. It constitutes an intrinsic injustice against him; an insult and a dishonour to him as a knower. Let us assume, also, that Vic regularly suffers identity-prejudicial credibility deficits: it is a persistent and systematic feature of his life. (Recall that ethnicity is a 'tracker' prejudice.) As a result he has suffered loss of confidence in some propositions: he holds lower credences than he otherwise would have in those propositions. Given that confidence is a prerequisite of knowledge (on many accounts), he has lost knowledge. Moreover, suppose that his intellectual development has been stunted; he has failed to develop the epistemic virtue of intellectual courage.

Given all of this, I think it is again reasonably clear what Ben is morally obliged to do. First and most clearly, he is obliged to try to counteract or nullify the intrinsic injustice: the insult and dishonour Vic

²² Not everyone share the intuition that Ben's obligations go so far as to eradicating the differences in benefits (and burdens) between him and Vic. However, nothing in what follows hinges on this. Others object, in various settings, 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. In this vein, 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 (Lawford-Smith 2016, pp.40–41). Similarly, Jeremy Dunham and Holly Lawford-Smith (Forthcoming) 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. In reply; *first*, I agree (as noted) that beneficiaries acquire obligations to seek to change the structures that generate injustice, but this does not indicate that beneficiaries do not *also* acquire obligations directly to the victims. *Second*, in the cases at issue in this paper (and others), it is clear that the recommended actions *would* benefit the victim in a just way. And *third*, as I have argued elsewhere (omitted), the perversion of incentives is in general a weak objection to a moral principle, because it focuses on only one part of what can and should motivate the relevant actors. Dunham and Lawford-Smith seem to suggest that ad-hoc redistributions remove an incentive those that sustain the status quo would otherwise have to reform it, namely the bad consequences for victims. But what should concern those that sustain discriminatory systems is not the position in which disadvantaged persons end up after intervention from individuals or civil society, but the position they would have ended up in without such intervention.

suffered in his capacity as a knower. Ben, having benefited from the injustice, is obliged to publically repudiate this in any way he can. He has a pro tanto obligation to do whatever he thinks will serve this end most effectively. This is likely to include publically criticising the committee's behaviour,²³ displaying that he values Vic's opinion, for example by soliciting his viewpoint regarding professional matters, listening attentively to him, and vigorously resisting any attempt from others to (again) undermine Vic as a knower, for example by interrupting him, or by trying to take credit for his insight or ideas.

As for the secondary epistemic harms, it seems clear that Ben is obliged to counteract the loss of knowledge and stunted intellectual development Vic suffers as a result of his continuing exposure to deficient credibility judgements. However, these stem *not only* from the injustice from which Ben is a beneficiary, but also from a long string of other injustices, so it's less clear how far Ben has to go here. The answer would seem to depend, 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 whites and blacks in the US. I will not pursue this question further here.

These obligations are clearly epistemic in nature. If the above is right, then, the Core Claim has been established: BFEI can give rise to moral obligations with a distinctively epistemic content.

It may be objected that *anyone* to whom this situation becomes transparent acquires the moral obligations I've attributed to Ben. As an unqualified statement, this is not plausible. It *is* plausible that anyone who's a member of a class of people who systematically benefit from injustices, including epistemic injustices, thereby acquire moral obligations (Thomson 1973). But these are more plausibly construed as obligations either to counteract such injustices *whenever they encounter them*; to counteract the system which sustains and enables these injustices in the most efficient way possible; or (most plausibly) both. For the vast majority of people in such groups this will involve no interaction with Vic, since the vast majority of the members of these groups will never encounter him, and the most efficient means of changing the relevant systems will not involve him.

A modified objection along these lines holds that anyone to whom the situation becomes transparent, and who is involved at least by being a bystander, acquires the kind of directed obligations which I have argued that Vic has. The idea is that one needs to be involved with the situation to obtain benefiting-sourced *directed* duties, and that, whereas the relation of being in a social group which routinely benefits does not qualify, the relation of being a bystander does.

²³ Thanks to NN6 for this point.

Suppose we accept this condition on acquiring directed duties, and that we accept the claim that bystanders acquire directed duties to the sufferer of injustice. This still does not imply that the moral obligations of bystanders to and beneficiaries of injustice are on a par.

A plausible way to operationalise differences in moral obligations is by reference to the cost people must accept to discharge them (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, and the costs she has to take on if she still does so are significant (Barry & Wiens 2014, pp.21–22).

This framework makes it easy to see the difference between beneficiaries and bystanders. Counteracting epistemic injustice in the way I have argued is morally obligatory for Ben entails costs for him. He may suffer damage to his standing or reputation in the firm, and to his further prospects for advancement, the probability that he will lose his job as a result of minor mistakes may increase, and so on. A bystander to the situation can appeal to quite a small cost along these lines to justify not discharging the obligation, but it seems clear that a beneficiary must appeal to a much higher cost. The involvement as a beneficiary of injustice increases the person's obligations significantly beyond that of a bystander, though of course not as far as that of a perpetrator of the injustice.

6. Congruence and Independence

I now turn to discussion and defence of the following two theses:

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

Let's again begin by making the claims clearer. Here I have in mind 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—perhaps they don't track natural kinds very well—but they will do for present purposes. The type of harm that has so far almost completely dominated discussions about BFI falls into the first category: a person is practically inconvenienced, or somehow left out of pocket. But a person can of course be unjustly harmed in many other ways: emotional abuse can cause harm and be unjust, for example, even if it has no practical consequences (perhaps because it is quickly and effectively counteracted).

What Congruence says is that, when the source of a person's moral obligation is that she innocently benefits from an injustice, it is the type of harm the victim has suffered—rather than, say, the type of benefit the beneficiary has gained, or any other factor—which *alone* determines the type of obligation the beneficiary accrues.²⁴

The second thesis says that epistemic obligations can arise in the absence of practical obligations, and practical obligations can arise in the absence of epistemic obligations. The latter half of the claim is uncontroversial; and the BFI literature is replete with examples. If Ariel steals Britt's car, and gives it Cory, Cory acquires practical obligations as a result (to relinquish the car), but no epistemic ones. The argument in §5 is intended to establish that epistemic obligations can arise as a consequence of a person benefiting from epistemic injustice; and I now treat that point as established. The question remains whether this can happen even if she acquires no practical obligations at the same time; is so, Independence is established.

So much for what the claims say. Should we believe them?

We have already encountered evidence for Congruence. For one, as noted, the BFI literature is replete with examples of victims of injustice experiencing practical harms, and those who argue that BFI gives rise to moral obligations almost without fail argue that the obligations in these cases are practical in nature. Moreover, it was only when we specifically introduced epistemic harm into the picture that epistemic obligations arose.

We can get further evidence by considering the plausible scenario in which Vic, in addition to the practical and epistemic harms, also suffers an emotional harm. As Congruence predicts, Ben thereby acquires emotional duties: moral duties with a specifically emotional content. He, more than a bystander, is obliged to do what he can to mitigate or reverse the emotional harm Vic has suffered.²⁵

We have seen evidence that type of harm bears on type of obligation. But Congruence says that type of harm *alone* dictates type of obligation, and one might easily think that both the type of harm *and* the type of benefit play a part here.

We obtain evidence against this view from reflection of a particular way of fleshing out the Two Accountants scenario. The case specifies that the candidates are visibly different in traits relevant to perceived ethnic identity. But it does not specify what some readers might have assumed, namely

²⁴ The *type* of obligation, not its strength, or where it falls on any other dimension of variation.

²⁵ He must of course do so in a sensible and sensitive manner, taking special care, for example, to avoid causing further harm by being patronizing, invasive, intrusive, and so on.

that exactly one of the applicants is white. Another scenario that fits the case is that both applicants are African Americans, but that they differ in their degree of similarity to the stereotype of African American people. There is strong evidence that prejudice against African Americans is stronger against those that are more similar to the stereotype (and evidence of the same phenomenon exists for other groups also, for example for Hispanics).²⁶

This opens up for the possibility of *both* candidates suffering identity-prejudicial credibility deficit in the Two Accountants case, but Vic more severely than Ben. It seems clear that this does not change the verdict; as the beneficiary (in practical terms) of injustice, Ben *still acquires* the epistemic obligations outlined above. He must still counteract the primary and secondary epistemic harms Vic has suffered to the best of his ability, and he must still take on a higher cost than a mere bystander in so doing.²⁷ So Ben acquires epistemic obligations by benefiting from injustice *even when* he does not benefit but instead is harmed epistemically, so long as he benefits somehow. The same is true, *mutatis mutandis*, for the emotional harm case. Together, these data-points lend significant credence to Congruence.

Let us turn now to the defence of Independence. Assuming that the argument in §5 is accepted, in order to establish this thesis it suffices to show that the epistemic obligations in the Two Accountants case remain when practical obligations fall away. To see this, consider the following amendment:

Career Change: As in Two Accountants, but shortly after being hired in the junior position, Vic is offered a position in middle management. Vic is indifferent between management and accounting as career paths, and the new job offers the same wage as the senior accounting role, as well as equivalent conditions for professional development.

²⁶ 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 predicting more skin tone bias among blacks than among whites, see (Linville et al. 1989). However, see (Maddox & Gray 2002) for evidence that both black and white people use skin-tone to socially categorize people they observe—skin tone is not, after all, masked by out-group homogeneity—and for evidence of the association of dark skin with negative traits and the converse: "Generally, dark skin tone was more often associated with poverty, aggressiveness, lack of intelligence, lack of education, and unattractiveness (p. 258)". For evidence from the IATs, see e.g. (Nosek et al. 2007).

²⁷ Naturally, if Ben has suffered epistemic injustice, other agents will have acquired epistemic obligations to him. Indeed, Vic may be among them in his capacity as a bystander (he is not a beneficiary). If he is, his obligations are also weaker than those Ben has acquired.

In this case it seems clear that Ben no longer has any practical obligations to Vic; since the practical harm Vic suffered as a result of the injustice has been entirely mitigated. But it also seems clear that the epistemic obligations remain. So, epistemic obligations can arise in the absence of practical obligations, and vice versa, and Independence holds.

7. Learning General Lessons about BFI from the Epistemic Case

The previous section illustrates that we can learn valuable general lessons about the correct theory of BFI from careful consideration of the special case of epistemic injustice, which is the final and most general claim this paper attempts to establish. This section pursues illustrates this further.

It was argued above that in Two Accountants, Ben is obliged to try to counteract or nullify the intrinsic injustice Vic suffers: the insult and dishonour to him in his capacity as a knower. These obligations pertain to the intrinsic injustice that obtains simply in virtue of Vic suffering a credibility deficit, regardless of its consequences. Reflection on this feature of the case is particularly illuminating with respect to the theory of BFI more generally. In the BFI literature it is sometimes held that benefiting from injustice 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 falls primarily on the perpetrator, and only secondarily and derivatively on the beneficiary.²⁸ A corollary of this is that, in the cases of practical harm which have so far dominated the literature, the beneficiaries' obligations can cease to exist if someone else—the perpetrator, or perhaps a collective insurance scheme—fully compensates the victim.

But 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 & Barry 2014); b) in the practical / monetary case, being made whole again *just means* being fully compensated, because money and (generally) property are *fungibles*—and c) that the perpetrator's obligation is *stronger* than the beneficiary's, in that the perpetrator must accept a higher cost to discharge it.

We can begin to see how we have been misled by focusing on possessions that are *not* fungibles. Suppose that one of Alice's most treasured possessions is a cigar. Aside from the significance it holds for Alice, there's nothing special about it: it is only moderately expensive and widely available. However, Alice's father, who recently succumbed to cancer, bought it 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

²⁸ See e.g. (Huseby 2013, p.3). Thanks to NN1 for discussion.

visiting Alice he realises that he's has forgotten to get his husband Chad an anniversary gift. Chad loves cigars. Bob sees Alice's cigar, and, planning to replace it with an identical one the next day, he takes it, wraps it up, and gives it to Chad, who 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, and she simply cannot be made whole again there, at least not in the short term. Nevertheless, both Bob and Chad are obliged to try as best they can: they must do whatever they reasonably think has the best chance of alleviating her hurt. And there's clearly nothing derivative about Chad's obligation here.

There are, of course, many cases of emotional harm where the victim of the injustice can be made whole—apologies often work. But in the case of epistemic injustice there seems to be nothing the perpetrator can do which will relieve the beneficiary from her obligation. Plausibly, the only way a victim of epistemic injustice can be made whole again is if *everyone* involved in the injustice—perpetrators, bystanders, *and* beneficiaries—publically repudiates the message of dehumanisation in a relevant way, for example by explicitly condemning the injustice and by explicitly honouring the victim as an epistemic agent. This difference between the epistemic and the practical cases traces, I think, to the fact that the primary harm is intrinsic—a feature epistemic harms share with emotional ones. So reflection on the cases of benefiting from epistemic injustice yields a more general insight about BFI: while beneficiaries' obligations may (or may not) be secondary and derivative in the cases of extrinsic harms, they are not where intrinsic harms are concerned. Indeed, although closer investigation is warranted here, the simpler and so *prima facie* preferable theory says that beneficiaries' obligations are *not ever* derivative or secondary, although they are weaker than those of perpetrators.

Let us turn now to the two suggested accounts of the demarcation criterion between cases of BFI that do, and those that do not, give rise to moral obligations. Haydar and Øverland suggested that at least one of three further conditions would have to be present: a structured competition that is perverted, that the perpetrator be motivated by a desire to benefit the benefactor, or that assets be transferred from perpetrator to beneficiary.

I think the epistemic case shows that these cannot, after all, be necessary conditions for BFI to generate moral obligations. It is true that the case examined in §5 does fit this account, since the epistemic injustice in that case does pervert a structured, and otherwise fair competition. But it is implausible that all cases are covered by this, or one of the other, criteria.

For example, take a case of three friends discussing the likely outcome of an upcoming election. If B and C disagree about a point where judgement clearly comes in, and A gives B's claim more credence than C's because of the operation of an identity-prejudicial stereotype against a group to which he takes C to belong (say, the gender female), clearly the first condition is not in operation. There is also no transfer of assets, nor is there a reason to think that A did this with the intention of benefiting B. (In fact, if conscious deliberation had been involved, this may have been sufficient to stop it from happening; we may suppose that A is, on the conscious level, vehemently opposed to the view that women are less likely to be knowledgeable about politics.) Yet it is easy to flesh out the case so that B benefits, and, as a result, he obtains more stringent obligations to counteract the harm which C has suffered than does a mere bystander (D, who's sitting nearby, not participating).

Turning now to the proposal due to Barry and Wiens, I should first note that it's not entirely clear to me whether these authors intend the kind of obligation I have been discussing to fall within the purview of their account.²⁹ However, if it does so fall, it seems to constitute a counterexample to that theory as well. Their core idea is that BFI gives rise to obligations only when the beneficiary's retention of the benefit *sustains the harm* suffered by the victim. On the account, there is a tight and substantial link between the two:

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).³⁰

²⁹ For example, they explicitly 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. Second, their restriction to cases of remedial duties is 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). Finally, they are concerned with duties to the victims, not with cases when the victim has passed away, a condition obviously met here.

³⁰ On the other hand, the opposite view, that the relation between the two needn't be substantive, and that the mere fact that both obtain suffices, may seem to be supported by the passage on p. 14, cited in §4 above.

Note the ‘by’—again indicating a substantial connection between the retention of the benefit, and the sustainment of the harm. In *Two Accountants*, however, there is no relationship between the benefit Ben received—money, and time for professional development—and the harm to Vic: the insult and dishonour he suffered in his capacity as a knower.

If the above is correct we have learned, from reflection on BFEI, in addition to the ones mentioned in earlier sections, the following general lessons about BFI. First, beneficiaries’ obligations are not always secondary to, or derivative upon, perpetrators’ obligations. Notably they’re not in cases of intrinsic harms, and the simplest theory says that they never are. Second, we have still not arrived at a generally applicable criterion of demarcation between the cases in which BFI gives rise to a moral obligation and those in which it does not, since BFEI provides counterexamples to both major criteria on offer. Again, these are lessons about how BFI should best be theorised, and are therefore independent of, and indeed strictly prior to, the question of whether BFI ultimately *is* an independent source of moral obligations. And the insights stem from reflection of the epistemic case in particular.

8. The Significance of Benefiting from Epistemic Injustice

It is easy to think that for epistemic injustice, the standard case is that all parties lose, and that cases of benefit are contrived and rare.³¹ After all, in cases of epistemic injustice, not only does *S* suffer epistemically, but *H* *does too*; by deflating their credibility judgement of *S* in spite of the evidence, *H* loses out on a source of knowledge he otherwise would have had. This notwithstanding, I think regarding BFEI as a fringe phenomenon is a mistake, for two reasons. First, injustice is the norm, not the exception. Second, a significant proportion of our speech is strategic.

Some might take the view that, at least in liberal democracies, justice is the norm, and injustice the aberrant exception. But the opposite perspective seems 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, p.39). Thus, for example, 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*

³¹ Thanks to NN2 for pressing this worry.

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, p.17).

The United States obviously has important features relevant to justice and injustice that are not shared with all other liberal democracies, notably a recent history of slavery. Still, I think the lesson generalises.

Considering specifically the case of epistemic injustice, it is clear that many of us would never countenance the conscious operation of identity-prejudice. That is not who we are; not who we allow ourselves to be. But Fricker rightly puts weight on the idea that such prejudice can do its dirty work ‘by stealth’, without our knowing, and even despite our consciously held beliefs (p. 39).³² And there is by now overwhelming evidence of the widespread existence and influence of implicit stereotypes across society, notably *also* from within the target groups themselves, for example, bias among blacks against blacks, or among women against women (Nosek et al. 2007). Given that we *very often* are in situations of testimonial exchange in which we do not have specific information about the participant, and given the prevalence of stereotypes (many of which are negative), the most plausible conclusion is surely, as Fricker also argues, that *epistemic injustice is everywhere*.

This does not yet entail that *benefiting* from epistemic injustice is widespread. However, I think we can show that this is likely.

In the course of arguing that Fricker overestimates the importance of identity prejudice for credibility deflation, Linda Martín Alcoff notes that Fricker allows that in certain situations, such as in her example of an honest car salesperson suffering a credibility deficit due to a stereotype, “structural background conditions generally preclude either honesty or gullibility ..., circumventing the possibility of authentic communication” (Alcoff 2010, p.132).³³ Alcoff goes on to argue that *strategic speech*—speech the purpose of which is not only, or not primarily, to transmit information and impart knowledge, but to a significant extent to further some agenda (to sell a product, to increase the likelihood of (re-)election, to further the interests of an institution, etc.)—in fact dominates:

³² See again n. 9.

³³ Fricker’s reply is in her (2010).

the excellent work on testimony beginning to develop in analytic philosophy runs the risk of irrelevance to real social life once again unless we find ways to address the *majority* of communicative interactions in which sincerity and trustworthiness are affected by the strategic designs of speech ... (p. 133).

Alcoff argues persuasively that in many instances, credibility judgements are *duly* deflated, because of a healthy scepticism to the speaker's agenda. However, it does not follow from this that credibility judgements aren't, in addition, *unduly* deflated further, by the operation of identity prejudices. So testimonial injustice can be present all the same, and our reasons for believing that they are widespread are unaffected by Alcoff's reasoning.

What Alcoff's point about the persuasiveness of strategic speech allows us to notice, however, is that speech very often forms part of a competition, for status, resources, and so on. When that is the case and someone's not believed, they (are more likely to) lose. And *then someone else wins*. In the many instances where TI is involved, those others benefit from epistemic injustice. For those concerned with justice, therefore, BFEI is not a marginal and esoteric phenomenon, but a central and important one.

9. Concluding Remarks

Over the last fifteen years or so, there has been an explosion of interest in whether, to what extent, and in which cases, benefiting from injustice is a source of moral obligation. Not surprisingly, attention has generally been focused on cases where the harm or injustice, on the one hand, and the benefit, on the other, are both practical. I have argued that, in some cases where an agent benefits from injustice of a different kind, namely epistemic injustice, she thereby acquires moral obligations with a specifically epistemic content. My broader aim has been to demonstrate by example that widening our view, and focusing on benefits and injustices of *all* kinds, allows us to better understand how to theorise BFI as a (potential) source of moral obligation. What, then, have we learned?

The discussion has shown that innocently benefiting from injustice sometimes gives rise to a directed obligation, on the part of the beneficiary, to accept a higher cost than a bystander must accept in order to make the victim whole again, *in whatever dimension(s) she was harmed* by the injustice.

This new theory of BFI commands our assent by its explanatory power. First, it explains the Core Claim, namely that innocently benefiting from epistemic injustice can generate moral obligations

with distinctively epistemic content. Benefiting results in such obligations precisely because epistemic injustice occurs when epistemic harm has been perpetrated.³⁴

Second, it explains Independence, the claim that epistemic obligations can arise in the absence of practical obligations, and vice versa. They can so arise since it is possible to (unjustly) harm someone in one of these dimensions without (unjustly) harming them in the other, and because the type of obligation depends only on the type of harm suffered.

Third, it explains why the obligations BFI give rise to are not in general derivative or secondary to the obligations of the perpetrator(s). We were misled to believe this by the focus on extrinsic dimensions of harm, chief among them practical (including monetary) harm, in which actors other than the beneficiary can succeed in making a victim whole again. The case of epistemic injustice, with its intrinsic aspect, allows us to see that making a victim whole again sometimes cannot be accomplished without the involvement of the beneficiary of the injustice, in which case there is nothing derivative about her moral obligation.

Finally, the discussion has provided support for a principled limit to beneficiaries' obligations: they are morally obliged to accept cost in order to discharge obligations stemming from having innocently benefited from an injustice only up until *a change of positions*.³⁵ This principle, too, is justified by its ability to explain our intuitions. Most obviously it explains the widely accepted upper limit in cases of practical harm, where most hold that the obligation is limited by the accrued benefits.³⁶ It explains this because money and possessions are *fungibles* with relatively stable value, so if a person were required to pay a larger sum of money than that which they had originally received unjustly, they would usually prefer not to have benefited to begin with. It also elegantly explains *exceptions* to this rule, even in monetary cases: an agent can fairly object to repaying the very sum she originally received if that sum (adjusted for inflation) now has a much higher utility for her, due to unfortunate

³⁴ Note that BFEI only *contingently* generates obligations with a distinctively epistemic content. After all, psychology and society might have been different than they actually are, and were they different enough, making a victim whole again in the epistemic dimensions might not have necessitated performing actions specifically addressing a person in her capacity as a knower (perhaps a song and dance would be best). As things actually stand, however, that is what it necessitates. Thanks to NN3 for this point.

³⁵ This principle seems to be what Daniel Butt has in mind (2007, pp.140–143), and is also, as noted, discussed favourably by Christian Barry and David Wiens (2014, pp.15–16).

³⁶ Avia Pasternak argues, I believe unsuccessfully, that obligations are *not* in this way limited in her (2014).

changes in her economic circumstances. Perhaps she would have preferred not having to repay even a quarter of the amount she originally received, since money was worth so much less to her then.

This limiting principle also works in the epistemic (and emotional) cases. As we have noted, at some point the costs of discharging one's moral obligations becomes so high that one would have preferred not to have benefited to begin with. At this point, but not before, it seems right to say that obligations end.³⁷ So the principle of change of position gives an account of the limits of moral obligations arising from BFI that is both principled and generally applicable.

³⁷ Presumably, this principle would need to somehow account for *wholly unreasonable* preferences; I leave this for another occasion.

10. References

- Alcoff, L.M., 2010. Epistemic Identities. *Episteme*, 7(2), pp.128–137.
- Anwander, N., 2005. Contributing and Benefiting: Two Grounds for Duties to the Victims of Injustice. *Ethics & International Affairs*, 19(1), pp.39–45.
- Arpaly, N., 2000. *Unprincipled Virtue: An Inquiry Into Moral Agency*, Oxford University Press.
- Barry, C. & Wiens, D., 2014. Benefiting from Wrongdoing and Sustaining Wrongful Harm. *Journal of Moral Philosophy*, pp.1–23.
- Butt, D., 2007. On Benefiting from Injustice. *Canadian Journal of Philosophy*, 37(1), pp.129–152.
- Cooley, E., Payne, B.K. & Phillips, K.J., 2013. Implicit Bias and the Illusion of Conscious Ill Will. *Social Psychological and Personality Science*, 5(4), pp.500–507. Available at: <http://spp.sagepub.com/cgi/doi/10.1177/1948550613506123>.
- Dovidio, J.F. & Gaertner, S.L., 2000. Aversive Racism and Selection Decisions: 1989 and 1999. *Psychological Science*, 11(4), pp.315–319.
- Dunham, J. & Lawford-Smith, H., Offsetting Race Privilege. *Journal of Ethics and Social Philosophy*.
- Fricker, M., 2007. *Epistemic Injustice*, Oxford University Press.
- Fricker, M., 2010. Replies to Alcoff, Goldberg, and Hookway on Epistemic Injustice. *Episteme*, 7(2), pp.164–178.
- Gawronski, B. et al., 2010. Attentional Influences on Affective Priming: Does Categorisation Influence Spontaneous Evaluations of Multiply Categorisable Objects? *Cognition & Emotion*, 24, pp.1008–1025. Available at: <http://www.tandfonline.com/doi/abs/10.1080/02699930903112712>.
- Goldman, A.I., 1999. *Knowledge in a Social World*, Oxford University Press.
- Goodin, R.E. & Barry, C., 2014. Benefiting from the Wrongdoing of Others. *Journal of Applied Philosophy*, 31(4), pp.363–376.
- Greenwald, A.G. et al., 2009. Understanding and Using the Implicit Association Test III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology*, 97(1), pp.17–41.

- Greenwald, A.G., McGhee, D.E. & Schwartz, J.L.K., 1998. Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), pp.1464–1480.
- Hahn, A. et al., 2014. Awareness of Implicit Attitudes. *Journal of Experimental Psychology: General*, 143(3), pp.1369–1392.
- Haydar, B., 2003. The Moral Relevance of Cost. *Philosophical Studies*, 112(2), pp.127–134.
- Haydar, B. & Øverland, G., 2014. The normative implications of benefiting from injustice. *Journal of Applied Philosophy*, 31(4), pp.349–362.
- Holroyd, J., 2012. Responsibility for implicit bias. *Journal of Social Philosophy*, 43(3), pp.274–306.
- Huseby, R., 2013. Should the beneficiaries pay? *Politics, Philosophy & Economics*, pp.1–17.
- Jones, K., 2002. The Politics of Credibility. In *A Mind of One's Own: Feminist Essays on Reason and Objectivity*. pp. 154–176.
- Jost, J.T. et al., 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, pp.39–69.
- Kaplan, M., 1985. It's Not What You Know that Counts. *Journal of Philosophy*, 82(7), pp.350–363.
- Langton, R., 2010. [Review of the book *Epistemic Injustice: Power and the Ethics of Knowing*, by Miranda Fricker]. *Hypatia*, 25, pp.459–464.
- Lawford-Smith, H., 2016. Offsetting Class Privilege. *Journal of Practical Ethics*, 4(1), pp.23–51.
- Linville, P.W., Fischer, G.W. & Salovey, P., 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), pp.165–188.
- Maddox, K.B., 2004. Perspectives on Racial Phenotypicality Bias. *Personality and Social Psychology Review*, 8(4), pp.383–401.
- Maddox, K.B. & Gray, S. a., 2002. Cognitive Representations of Black Americans: Reexploring the Role of Skin Tone. *Personality and Social Psychology Bulletin*, 28(2), pp.250–259.

- Madva, A., 2017. What can we learn from the Implicit Association Test? A Brains Blog Roundtable. *The Brains Blog*.
- Maitra, I., 2010. The Nature of Epistemic Injustice. *Philosophical Books*, 51(4), pp.195–211.
- Miller, D., 2001. Distributing Responsibilities. *Journal of Political Philosophy*, 9(4), pp.453–471.
- Mills, C.W., 2007. White Ignorance. In *Race and Epistemologies of Ignorance*. pp. 11–38.
- Nosek, B. a. et al., 2007. Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, 18(1), pp.36–88.
- Oswald, F.L. et al., 2013. Predicting Ethnic and Racial Discrimination: A Meta-Analysis of IAT Criterion Studies. *Journal of personality and social psychology*, 105(2), pp.171–92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23773046>.
- Pasternak, A., 2014. Voluntary benefits from wrongdoing. *Journal of Applied Philosophy*, 31(4), pp.377–391.
- 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), pp.181–192.
- Payne, K. & Lundberg, K., 2014. The Affect Misattribution Procedure: Ten Years of Evidence on Reliability, Validity, and Mechanisms. *Social and Personality Psychology Compass*, 8(12), pp.672–686.
- Singer, P., 1972. Famine, Affluence, and Morality. *Philosophy and Public Affairs*, 1(3), pp.229–243.
- Smith, C.T. & Nosek, B. a., 2010. Implicit Association Test. In *Corsini's Encyclopedia of Psychology*. Wiley, pp. 1–4.
- Steinpreis, R.E., Anders, K.A. & Ritzke, D., 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), pp.509–528.
- Stroud, S., 2006. Epistemic Partiality in Friendship. *Ethics*, 116(3), pp.498–524.
- Tang, W.H., 2009. *Belief and Credence*. The Australian National University. Available at: <http://whtang.com/beliefandcredence.pdf>.
- Thomson, J.J., 1973. Preferential Hiring. *Philosophy & Public Affairs*, 2(4), pp.364–384.

