



Explanatory Injustice and Epistemic Agency

Veli Mitova¹ 

Accepted: 24 May 2020 / Published online: 3 June 2020
© Springer Nature B.V. 2020

Abstract

What is going on when we explain someone's belief by appeal to stereotypes associated with her gender, sexuality, race, or class? In this paper I try to motivate two claims. First, such explanations involve an overlooked form of epistemic injustice, which I call 'explanatory injustice'. Second, the language of reasons helps us shed light on the ways in which such injustice wrongs the victim *qua* epistemic agent. In particular, explanatory injustice is best understood as occurring in explanations of belief through a so-called reason-why when the correct explanation in fact features a motivating reason. I reach this conclusion by arguing that such explanations are a kind of normative inversion of confabulation. Thinking in these terms helps us see both how certain reason-ascriptions empower while others disempower, and (consequently) how through them believers are robbed of agency over their beliefs.

Keywords Epistemic injustice · Epistemic agency · Epistemic reasons · Rationalisation · Confabulation · Reason-why · Motivating reason

1 Introduction

What is going on when we explain someone's beliefs by appeal to stereotypes associated with her gender, sexuality, race, or class, instead of the reasoning that lead her to believe as she does? Suppose Lindiwe complains to Joe that he has never been supportive enough in her career. Suppose, further, that she is right to believe this, but he shrugs her off by saying that she is just being emotional and overly sensitive. I think most of us would agree that Joe's explanation of Lindiwe's belief is unjust in some way, and that such injustice is the bread and butter of sexist, racist, and other X-ist discourse. This paper tries to motivate two further claims: this is a distinctive form of epistemic injustice; and thinking of it in the language of reasons illuminates the relationship between epistemic injustice and agency.

✉ Veli Mitova
vmitova@uj.ac.za

¹ Department of Philosophy, University of Johannesburg, PO Box 524, Johannesburg 2006, South Africa

The phrase ‘epistemic injustice’ covers any unjust epistemic relation which disadvantages someone in her capacity as knower. The two paradigmatic instances, with which Miranda Fricker (2007) introduced the concept, are testimonial and hermeneutical injustice. Testimonial injustice occurs when a speaker’s testimony is given less credibility than it normally would, due to identity prejudice on the part of the speaker. Fricker (2007: 23) uses the example of a black man in the ‘30s American South being wrongfully accused of raping a white woman. His testimony during the trial is, predictably, discredited and twisted by the white jury, clearly due to racism and the race power dynamics of the time.

Hermeneutical injustice is a more structural phenomenon.¹ It occurs when a person lacks the epistemic resources that would enable her to make sense of her experiences or communicate them to others. The classic example here is when we didn’t have the concept of ‘sexual harassment’. Without this concept, victims of such harassment were unable to fully understand their experience or to resist its harms and justify their actions to others by appeal to it (Fricker 2007: 147).

Since the publication of Fricker’s monograph, many other kinds of epistemic injustice have emerged in the literature (e.g., Dotson 2011; Peet 2017; Pohlhaus 2012). The first aim of this paper is to spotlight a new kind – ‘explanatory injustice’.²

The second aim is to show that the language of reasons (rather than the more usual one of virtue) illuminates the ways in which explanatory injustice involves wronging the victim in her capacity as epistemic agent.³ In particular, I argue, such injustice occurs when we explain a belief through a so-called reason-why when the correct explanation in fact features a motivating reason. I reach this conclusion by arguing that such explanations are a kind of normative inversion of confabulation. Thinking in these terms helps us see how certain reason-ascriptions empower while others disempower, and (consequently) how through them believers can be belittled as epistemic agents.

I proceed as follows. First, I say more about the distinction between motivating reasons and reasons-why (§ 1). I then show how confabulation involves the self-attribution of the former when, in fact, only the latter are present (§ 2). Next, I argue that one fundamental kind of X-ist explanation of action involves the inverse normative distortion (§ 3). Fourth, I show that a similar inversion is present in *belief*-confabulation and X-ist explanations of *belief* (§ 4). Fifth, I situate the resulting explanatory injustice on the map of more familiar forms of epistemic injustice (§ 5). I conclude by addressing two potential worries (§ 6).

2 Kinds of Reasons

In this section, I distinguish the kinds of reasons we will need for understanding explanatory epistemic injustice.

¹ But see Dotson (2012) for an argument that we shouldn’t understand it *just* as a structural phenomenon.

² One might think that explanatory injustice is a form of what Andrew Peet (2017) calls ‘interpretative injustice’. A paradigmatic case is when a man hears a woman’s ‘no’ as a ‘yes’, in line with the stereotype of women liking to play hard to get. But this kind of injustice is different in that its vehicle is the interpretation of utterances, rather than the explanation of belief.

³ I use ‘wrong’ rather than ‘harm’ deliberately. If I have a racist belief about you when I am permanently stuck on a desert island, I can never harm you, but I am clearly wronging you.

2.1 Reasons for Action

Suppose you are reading in the lounge, and you hear Lindiwe and Joe fighting next door. Tempers are rising; voices are raised. Then there is silence, and the street door slams. Joe comes into the lounge red-eyed.

Now imagine two dialogues.

DRAFT

You: Did Lindiwe slam the street door?

Joe: No. She's still next door.

You: Strange. Then why did the door slam?

Joe: Probably the draft banged it shut.

INDIGNATION

You: Did Lindiwe slam the street door?

Joe: Yes.

You: Why?

Joe: Probably because I told her I'd cheated on her.

Two differences between the dialogues are relevant here. First, in *DRAFT*, Lindiwe has not performed an action, while in *INDIGNATION* she has. In *DRAFT*, a mere event occurred: the *door* slammed; its cause was the draft.⁴ In *INDIGNATION*, by contrast, a much more special event occurred – an *agent* brought something about: Lindiwe slammed the door.⁵

The second difference between the two cases lies in the explanation of what happened. In both we have an answer to the question 'Why', but they are different kinds of answer. In *DRAFT* we have a *merely causal* explanation; in *INDIGNATION*, we are offered a *rationalising* one. A merely causal explanation is one that cites a brute cause in answer to the question why an event occurred. A rationalising explanation is one that cites a special kind of cause, typically of an action or belief, one that makes the action or belief intelligible.⁶

⁴ I won't pretend to have an account either of causal explanation or of the ontology of causal relata. These are complex issues (see, e.g., van Fraassen 1980; Woodward 2014), but I trust that we all have enough of an intuitive grip on the concept of cause to be able to distinguish a mere cause from a rationalising one.

⁵ Precisely how we want to characterise what happened here is a delicate matter. (See, e.g., Davidson 2001: Essay 3.) The present point is simply that there will be plenty of obvious descriptions under which an agent did something in *INDIGNATION*, but none in *DRAFT*.

⁶ Some philosophers deny that reasons are causes (e.g., Dancy 2000). For arguments that the denial is ontologically too costly, see Mitova (2017: § 3.2.2).

Philosophers of action and metaethicists mark this distinction by speaking of two kinds of reasons – a reason-why and a motivating reason.⁷ A reason-why is a plain old cause. A motivating reason is a consideration that the agent took to favour her action or belief, and that hence moved her to act or believe as she did. Thus, in *INDIGNATION*, Lindiwe took the fact that Joe cheated on her to favour slamming the door on him. Citing such reasons explains the action by making it intelligible.

Typically, the correct explanation of action features motivating reasons, the sort of considerations that both cause and rationalise the action.⁸ But some of our explanations feature reasons-why. Consider a third dialogue.

DRINK

You: Did Lindiwe slam the door?

Joe: Yes.

You: Why?

Joe: Because she is drunk, and she always gets violent when she drinks.

Like *DRAFT*, *DRINK* cites a reason-why for the happening. But unlike *DRAFT* and like *INDIGNATION* the happening is an action: an agent, wobbly as she was, did something.

Sometimes the explanation of an action in terms of a reason-why will be correct. It may be that there is simply nothing rationalising to be cited in its favour. At other times, the explanation won't be correct. Such cases will be the topic of § 3. For the moment, the important thing is this. One way of thinking about the distinction I am drawing here is that a reason-why cannot be taken by the agent to favour or make intelligible the action in question. Thus, while Lindiwe can slam the door because she is drunk, in the causal sense of 'because', she cannot coherently think to herself, 'I am drunk, so let me slam the door'. Being drunk simply doesn't make slamming the door intelligible. (Though it can be a motivating reason for *other* actions – like going to bed or taking a cold shower, say.) A motivating reason, by contrast, can be, and typically is taken by the agent to favour the action. Thus, Lindiwe *can* think to herself, 'Joe has cheated on me, so I will slam the door on him'.

Motivating reasons can feature in deliberation in this way, because, unlike reasons-why, they can be *good* reasons, or more technically, 'normative reasons'. These are considerations that in fact favour the action (though they needn't do so conclusively).⁹ Thus, when things go well, the agent's reason for doing something is in fact a normative reason for doing it. That someone wronged you by cheating on you in fact favours slamming the door on him. And

⁷ Examples include Alvarez (2010) and Dancy (2000). Some authors use 'explanatory reasons' instead of reason-why, but this label courts confusion. First, it misleadingly suggests that motivating reasons aren't explanatory. And this leads to various problems in the ontology of reasons debate. (I have argued for this in Mitova 2017: § 1.2). Second, as a matter of fact, the literature vacillates in its use of 'explanatory reason', partly because the term is ambiguous between purely causal explanation and a rationalising one. Here, for example, is Parfit using it as a synonym of a motivating reason, instead of a reason-why: 'When we have [a normative] reason, and we act for that reason, it becomes our motivating or explanatory reason' (Parfit 2001: 17).

⁸ 'Correct' here and throughout doesn't mean 'ideal', but simply 'faithful to what in fact moved the agent to act/believe as she did'. Thanks to an anonymous reviewer for requesting clarity here.

⁹ This 'favouring' analysis of normative reasons isn't without its opponents (e.g., Hieronymi 2005). See Mitova (2017: § 1.2.1) for why such opposition is misguided.

when Lindiwe takes it to do so, as she does in *INDIGNATION*, she acts for a good reason. But there is no such fit between a reason-why and a good reason: being drunk simply can't be a good reason for slamming the door.

2.2 Reasons for Belief

These distinctions amongst kinds of reasons can be equally drawn for belief. Consider the following pair of dialogues:

JEALOUS

Joe: Ugh, Lindiwe just walked out on me because she believes I've cheated on her.

You: Why does she believe that?

Joe: Because she is pathologically jealous.

EVIDENCE

Joe: Ugh, Lindiwe just walked out on me because she believes I've cheated on her.

You: Why does she believe that?

Joe: Because I told her I had.

We can say here pretty much the same things as we said about the difference between *DRINK* and *INDIGNATION*.¹⁰ In *EVIDENCE*, Lindiwe's belief is explained by appeal to her motivating reason, or her evidence – Joe's testimony, something she took to favour the belief that he has cheated on her. In more familiar epistemology-speak, a motivating reason is the reason on which Lindiwe based her belief.¹¹ This reason can be a good reason – a consideration that in fact favours believing Joe unfaithful. It is, hence, something that she can take up in deliberation about what to believe: 'Joe told me he cheated on me; so, he probably did.'

By contrast, no such motivating reason is available in *JEALOUS*. Only a reason-why is. The fact that you are pathologically jealous is not something that favours believing your partner unfaithful. Consider: 'I am pathologically jealous; so, he probably cheated on me'.

JEALOUS features a clear-cut case of a reason-why explanation of belief. Arguably, another instance of such explanation are cases when we cite a pragmatic reason for the belief. For instance, Lindiwe could have formed the belief above, say, because she was sick of Joe and was looking for an excuse to leave him. This is only a reason-why, the thinking would go, because it cannot feature in our explicit deliberation about what to believe.¹² Thus, while such explanations make the belief pragmatically intelligible, they don't make it *epistemically* so.¹³

¹⁰ I don't mean to suggest that the concepts of motivating reason and reason-why will have the same extensions in the action and belief case, just that their intension is the same.

¹¹ The so-called basing requirement, which many accept, is the requirement that in order for your belief to be (doxastically) justified, it needs to be held for the good reasons you have for it (e.g., Neta 2002).

¹² This is meant simply as a psychological claim that is accepted by pretty much everyone. (The *locus classicus* is Williams 1973.)

¹³ I am obviously skirting rather brashly the debate about whether pragmatic reasons can be normative reasons for belief. For arguments that they can be, see e.g., Reisner (2009).

3 Confabulation

I have introduced the distinction between reasons-why and motivating reasons at such length, because it can help us make perspicuous the power dynamics of reasons-ascriptions. I said that motivating reasons provide *rationalising* explanations by being both the cause of the action or belief and the thing that makes the action or belief intelligible. Since the notion of a cause is factive, this means that the agent must have acted for that reason.¹⁴

But there is another kind of rationalising explanation, with which we are familiar from everyday life – what psychologists call ‘rationalisation’ or ‘confabulation’. (I will stick to the latter term, to avoid confusion.) You have had too much coffee and irritably tell me off. When I suggest that it is not you but the coffee speaking, you indignantly fabricate on the spot a perfectly intelligible – but of course false – story about why I deserved to be told off.

Such confabulation, according to some researchers, is disturbingly pervasive. Let me give one case, based on a familiar experiment (Nisbett and Wilson 1977), so that we have something real to hold onto.

STOCKINGS

You and I, and a whole bunch of other people, are shown four identical pairs of stockings and asked to choose one. Most of us choose the one on the furthest right. When asked why we chose that one, no one cites the position of the stockings; each of us gives a reason that clearly shows the superiority of the choice we made – the chosen pair was silkier, finer, classier, you name it.

This case, along with many others, epitomise one kind of choice-confabulation. Some have argued that much of our reasoning in fact follows this and other irrational patterns. Others have disputed both this claim and the experimenters’ interpretation of the above results.¹⁵ But no one, to my knowledge, denies that confabulation sometimes happens. This is all my argument needs. So, I will use the above case merely as a handy illustration; the reader is welcome to replace it with her favourite example of confabulation.

Here are the three features of such cases that interest me. First, the sufficient cause of the choice (as far as reasons are concerned) is some bias.¹⁶ In *STOCKINGS*, it is a bias in favour of things on one’s right. Second, no true rationalising explanation of the choice is in the wings. (This feature is entailed by the first, since the cause was said to be sufficient.) In *STOCKINGS*, we are told that the stockings are identical, so there can’t be a good reason for choosing one pair over others. Given the large number of people who settle on the same pair, we can also be pretty sure that there isn’t a motivating reason at work, but it is made up in retrospect. Finally, the rationalising explanation that is in fact offered in *STOCKINGS* and other such cases doesn’t obtain. This feature is entailed by the second – and *a fortiori* by the first. Since I was caused by

¹⁴ Some philosophers deny that motivating reasons need to be the ones the agent actually acted for (e.g., Broome 1997: 88). For an argument why this is both linguistically and theoretically ill motivated, see Mitova (2017: § 1.2.2).

¹⁵ See, for instance, Newell and Shank (2014).

¹⁶ Clearly, many reasons-irrelevant things need to obtain in order for the choice to be made. (You need to have a head, for starters.) As John Mackie’s hunt for the INUS condition has taught us, causal explanation, especially of the kind we are after here, is a very tricky matter. (See Mackie 1965, and for an example of a critique Jackson 1982.) I have already owned that I don’t feel up to settling it here (note 3).

bias to choose the stockings, the reason I am citing as mine – their silkiness, say – is not what moved me to choose.

If these are genuine features of a certain kind of confabulation, then we can give the following schema for it:

*CONFABULATION**:

S makes a choice or performs an action for nothing other than a reason-why at t_1 , but self-ascribes an ostensible motivating reason at t_2 .

I will just focus on this kind of confabulation here: it essentially involves the self-promotion from the space of mere causes into the space of reasons and hence agency.

I have qualified it with an asterisk, so as not to imply that it is the only kind of confabulation out there. Another common kind arguably occurs in cases in which an action is performed out of a morally suspect *motivating* reason, but one cites to oneself a morally good motivating reason. (Say that I bought you chocolates to curry favour with you, but persuaded myself that I did it to cheer you up.) I suspect that this is a less empowering confabulation, but don't have the space to defend the suspicion here. The point of mentioning it is simply to distinguish it from my present target, which involves cases where the agent didn't act for *any* motivating reason, but self-ascribes one nonetheless.

Suppose that we¹⁷ indulge in this kind of confabulation because, as we keep being instructed, we are meant to be rational creatures, and lapses into irrationality unsettle us so much that we need to make up an intelligent story where none exists. I say 'suppose'. It seems to me that the research on cognitive dissonance (starting with Festinger 1956) supports this suggestion, but I am not concerned with a psychological account of these phenomena.¹⁸ So here is my claim: *if* the 'suppose' is true, then it would explain an interesting fact about the relative powers of motivating reasons and reasons-why. Citing motivating reasons makes the explanation look respectable and the actor look good. In contrast, citing reasons-why alone demeans the actor by offering a merely causal explanation of her action. Another way of putting it: motivating-reasons ascriptions empower; reason-why ascriptions disempower; this is why we are at such pains to confabulate good-looking explanations of our actions. Treat this as a hypothesis that may be useful in analysing some forms of epistemic injustice. (Whether it *is* useful, the reader can judge in § 4.)

Two thoughts lend some prima facie support to the claim that the one kind of reason empowers and the other disempowers, even if it turns out that I am wrong about why people engage in confabulation. First, many philosophers think that being done for a reason is a constitutive feature of action. Hence, Elizabeth Anscombe's (1976) famous characterisation of action as something to which a particular sense of the question 'why' is applicable. So, to merely cite a cause in the explanation of my action is somehow not to acknowledge that it *is* a full-blown action, and hence to deny me agency in it.

Second, we really are demeaned and feel like we are not being taken seriously by being explained away like this. Think of traditional putdowns such as 'She is only saying that because she is emotional', or 'He is only acting like this because he hasn't taken his meds'. If you are the she or he in question and realise that the explanation is correct, the realisation will

¹⁷ I am well aware of the barbed sound to this pronoun. One of the important lessons of recent work in experimental philosophy is that such intuitions can vary with culture (Weinberg et al. 2001) and gender (Buckwalter and Stich 2014).

¹⁸ More recently, some have argued that our need for consistency is, in fact, an integral part of moral progress (e.g., Campbell and Kumar 2012; Summers 2017).

come as a vexing surprise to you. If it is not correct, you will probably feel like you are not being given your due.¹⁹

Although these examples involve the agent's being aware of the demeaning explanation, my claim doesn't depend on such awareness. The subject of the explanation would be demeaned *qua* agent regardless of whether she is aware of it. It is enough that she is wrongfully represented as less than a full agent. I will return to this point in § 4.2.²⁰

These thoughts, of course, hardly provide deductive support for my claim that reasons-why disempower while motivating reasons empower. It is not as though because action is something done for a reason, citing a cause disempowers me. But I think they do lend the claim abductive credibility: the disempowering effect of reason-why ascriptions would be a good explanation of why I feel demeaned when explained in this way. It may be that these powers of reason-ascriptions are due to a particular androcentric, over-intellectualised view of 'man' as the rational animal. Be that as it may, many of us have internalised them to a greater or lesser degree, and hence accounts of reasons-ascriptions and of epistemic injustice which pretend they don't exist are incomplete.

This then is confabulation*: it involves the self-ascription of a supposed motivating reason when, in fact, only a reason-why is available.

4 Reasons for Action and Power

These morals were drawn from reflecting on self-ascriptions of reasons. But as feminist, race- and queer-theorists have taught us, power is often a zero-sum game. If someone's empowered in the social domain, it is typically at the expense of someone else (e.g., Mills 2007: 15). This is obscured by focusing on self-ascription and isolated subjects, as I have been doing so far. So, I now move to other-ascriptions of reasons, and argue that one of the most pernicious forms of sexist, racist, and any X-ist explanation of action is the exact inverse of confabulation*.

Remember Joe and Lindiwe from § 1.1? Let us carry on a bit from *INDIGNATION*.

GIRL

You: Why did Lindiwe slam the door?

Joe: Because I told her I'd cheated on her.

You: So?

Joe: Ugh, you know girls, all emotion and no thought.

Here is what just happened in the language of reasons: you and Joe took away Lindiwe's motivating reason (justly feeling wronged, or however you prefer phrasing it), and replaced it with a reason-why. You did it by not seeing as intelligible something which is perfectly obvious to anyone with the barest grasp of monogamous relationships. Joe added insult to

¹⁹ We can also use such phrases to *excuse* people. Although not disempowering, such excuses still help support the larger point here: they work only insofar as they deny the actor agency over her action.

²⁰ Thanks to an anonymous referee for helpfully pressing me on this point.

injury by spelling out what was driving your incomprehension in any case – the thought that there is no question of intelligible explanation when ‘girls’ are at issue.

I suggest the following schema for this kind of explanation (where ‘S’ stands for ‘subject’ and ‘E’ for ‘explainer’):

X-IST EXPLANATION OF ACTION*

The correct reason for S’s action is provided by a motivating reason, but E cites instead a reason-why in explaining S’s action.

Three points about this schema. First, it clearly oversimplifies by supposing that there is a single correct motivating reason for each action, which is moreover easily identifiable. I think that typically both of these assumptions would be correct, but those who feel uneasy are welcome to replace ‘correct reason’ with ‘correct *sufficient* reason’ or ‘*most salient* correct reason’. Nothing in my argument turns on leaving out these complications. I just do so for the sake of space.²¹

Second, the schema is most certainly not meant to offer a fully-fledged analysis of X-ist explanation. As it stands, it is compatible with the option of positive reasons-why making the explanation X-ist.²² Say that you bought me chocolates in order to cheer me up, but I explained it by saying ‘You just can’t help yourself; it’s your generosity’. Then I am ascribing to you a reason-why when you had a motivating reason, but clearly the explanation isn’t X-ist in any obvious way. So, a fully-fledged analysis would have to say what else needs to be in place in order for the explanation to be properly X-ist and to thereby constitute a wrong. I will return to this point in § 4.2. For now, the point of the schema is to provide one hopefully illuminating way of capturing the power-dynamics of reasons-ascriptions in this kind of explanation. This, in turn, will afford us the beginnings of an account of the way our epistemic agency is wronged in explanatory injustice.

Third, and starting work on this aim, it should be readily apparent that the above schema is the precise inverse of confabulation*, where something was done for a reason-why alone but a loftier explanation in terms of a motivating reason was offered. What is the significance of this normative inversion? I think this: just as in confabulation* the subject is trying to elevate her action and hence herself, by citing a proper reason when only a reason-why is available, so in X-ist* explanations the explainer is demeaning the explained by giving a mere reason-why where a proper reason is available. Since, as discussed in § 2, to deny someone a motivating reason is to deny her agency in her own action, this type of explanation is a fundamental form of disempowerment.²³

It is worth emphasising again (as I did in the confabulation case) that the sense in which Lindiwe is robbed of agential power doesn’t depend on her being aware that an X-ist explanation has been given of her action. Incorrectly being represented as lacking in agency wrongs the agent regardless of whether she knows about the wrong: in a world in which social identities are partly constructed by others, agential authority in part depends on others’ acknowledgement of it. When the acknowledgement is wrongfully withheld, our agency is belittled and an injustice is perpetrated.

²¹ Thanks to an anonymous reviewer for pressing me on this point.

²² Thanks to an anonymous reviewer for pushing me on this point in the context of the example in § 4.2.

²³ Which is not to say, of course, that the corresponding injustice is the only one worth combatting, nor that if we eradicated it, other injustices would disappear.

Finally, let me briefly contrast this kind of X-ist explanation with another, corresponding to the other type of confabulation briefly discussed in § 2. The vehicle here are actions done for a motivating reason, but explained in terms of a less morally good *motivating* reason. Think, for instance, of explaining a woman's action as due to women's apparently insatiable need to ensnare potential husbands. When such explanations are due to identity bias, large groups of people deprive other groups of people of their rightful *moral* authority. In X-ist* explanations, by contrast, the agent is deprived of *agential* authority simpliciter.

5 Epistemic Reasons and Power

So far, I have argued that one fundamentally disempowering form of biased explanation of action is the obverse of an empowering form of confabulation*. In the latter, we empower ourselves by incorrectly self-attributing a motivating reason; in the former, we disempower another by incorrectly attributing to her a reason-why. The talk of power is literal in both cases: we either endow ourselves with agency over our action or represent another as having no agency over hers.

What does all this have to do with epistemic injustice? The answer, of course, is that when you don't grant people their due *epistemic* agency, you are in the business of unjustly depriving them of epistemic power, hence inflicting an epistemic injustice on them.

Many will be unhappy about talk of epistemic agency and of purely epistemic powers. One way of looking at this paper is as an argument for legitimating such talk. Talk of epistemic agency is made more plausible by the structural analogies between the epistemic and practical normative domains which I highlight here. And talk of epistemic powers is legitimated in four steps. First, I have already shown that reasons-ascriptions are vehicles of power. Second, and in what follows, I develop the epistemic analogue of the relationship between confabulation* and X-ist* explanation (§ 4.1). Third, I argue that these latter phenomena involve similar power exchanges as their action counterparts (§ 4.2). Finally, I point to some common instances of such exchanges in order to make sure of the reality of the injustice I am trying to isolate (§ 4.3).

5.1 Epistemic Confabulation

Recall *STOCKINGS*, in which people displayed a bias in favour of things on their right, but would instead explain their choice by citing all sorts of reasonable things like the greater softness of the stockings. These kinds of cases, I suggested, have the following shape:

*CONFABULATION**:

S makes a choice or performs an action for nothing other than a reason-why at t_1 , but self-ascribes an ostensible motivating reason at t_2 .

Although this schema characterises an action- or choice-confabulation, we can easily replicate it for belief. For the confabulation of the choice is in fact achieved through the confabulation of a belief that is constitutive of the choice. The content of the belief in this case is something like this:

BEST Pair 4 is superior to the other three pairs.

What made me believe *BEST* was something like this:

RIGHT-BIAS When asked to choose amongst identical objects, most agents go for objects on their furthest right.

But what I cite in confabulation is:

SOFT Pair 4 is the softest.

SOFT would be a good reason for believing *BEST*. But, the experiment teaches us, *SOFT* is not in fact the agent's reason – or the reason on which her belief is based. Rather, the belief was caused by *RIGHT-BIAS*.

We can then have the following characterisation of this type of belief confabulation:

*BELIEF CONFABULATION**

S adopts a belief at time t_1 for a reason-why, but self-ascribes an ostensible motivating reason at t_2 .

And we can tell the same story as we did in the action case. Suppose we treated as a hypothesis the claim that we belief-confabulate because, given certain historical contingencies and centricities, we are distressed by thinking of ourselves as epistemically arational. So, we just make up a motivating-reason story for why we believe as we do. This hypothesis would then explain certain power asymmetries in citing one explanation of the belief over the other. Thus, the true, *RIGHT-BIAS* explanation of my belief in *BEST* is disempowering. It makes me look theoretically arational (for the good reason that I am). So, I empower myself by self-ascribing a motivating reason instead, citing *SOFT*. A motivating reason, in this context, would be a consideration that favours believing the proposition in question. Typically, this is something that the agent takes as evidence of the truth of that proposition.

Notice that the sort of power involved here is distinctly epistemic. It is not, in this case, that I am empowering myself as the agent of an action or choice, but rather by construing myself as *believing* for the right *epistemic* considerations. That is, I ascribe myself epistemic authority by representing my belief as epistemically rational (or justified, or epistemically good in some other way). This is a purely epistemic power-gain, even though, the resultant *choice*-confabulation also empowers me through endowing me with practical agency where none was present before.

If this is on the right track, epistemic motivating reasons are shown to come with epistemic powers. We should thus expect that at least some forms of epistemic injustice would involve the sort of power-inversion of confabulation* that we found in the practical case (§ 3.2).

5.2 Epistemic Injustice and Reasons

Let's consider a last (you'll be heartbroken to hear) dialogue:

MATHS STAR

You: Nkanyezi thinks he will be a great mathematician.

Joe: Ha! Wishful thinking. We all know that black people aren't strong on abstract thought.

You: But he says he is already at the top of his third-year class. And he did the entire degree in one year, even though he is only fourteen!

Joe: Yeah, well, the lecturers are soft with him because he is a teacher's pet.

MATHS STAR is a typical instance of biased explanation of belief. According to Nkanyezi himself, his belief that he will be a great mathematician is based on what by any standard is excellent evidence. Most of us find *school* maths pretty tricky at the age of fourteen, let alone taking a whole university degree of the stuff in one year. So, he clearly has a motivating reason – and a good one to boot – for believing as he does. Joe's explanation of the belief, however, features a mere reason-why – wishful thought. When the facts are brought to Joe's attention, he immediately discounts them as proper reasons by trying to show that they don't in fact favour Nkanyezi's belief.

As in the action case, a mere reason-why explanation will sometimes be appropriate for belief. Think of the proverbial cuckolded husband who holds onto the belief that his wife is faithful in order to preserve his self-esteem. Arguably, Joe's own belief that Nkanyezi is indulging in wishful thought is of this kind, too. But in Nkanyezi's case, such an explanation is plainly inappropriate. It can only be due to recalcitrant bias, as the dialogue should make obvious.

Here is a schema for such explanations:

X-IST EXPLANATION OF BELIEF*

S holds a certain belief for a motivating reason (i.e., the belief is based on good evidence as far as *S* is concerned); but *E* cites instead a reason-why in explaining *S*'s belief.

This way of putting things hopefully makes clear that this explanation exactly mirrors Belief confabulation*, and hence the disempowerment attendant on citing a reason-why in place of the real motivating one. In case it isn't quite clear: in confabulation, the rationaliser epistemically empowers herself by construing her belief as epistemically rational. This is done by citing an epistemic motivating reason in place of the real reason-why. In X-ist* explanation, conversely, the explainer epistemically disempowers the explained by construing his perfectly rational belief as irrational, or worse arational.²⁴ This is done by citing a mere reason-why in place of the real motivating reason.

Notice that, as with its action equivalent, the schema is not meant to provide a fully-fledged specification of X-ist* explanation. In particular, it leaves it open that incorrect *positive* reason-why explanations also count as X-ist*. This might sometimes be true (e.g., with appeals to the stereotype of Asian students being great at maths), but typically it won't be. I think that the most plausible fix here would be to add the condition that the reason-why ascription (positive or negative) be due to identity prejudice.²⁵ This would give us a more comprehensive account of the wrongs involved in X-ist* explanations, but I don't have the space to defend it here. It will be enough if the schema is taken to isolate a perspicuous connection between reasons-ascriptions and agency.

²⁴ I am being deliberately non-committal here. My intuition is that 'arational' is the right label whenever a reason-why is the sole explanation. But staking this claim for belief would, I think, commit me to a side in the evidentialism-pragmatism debate, of which I promised to steer clear (see note 13).

²⁵ Thanks to an anonymous referee for the suggestion.

I have told the epistemic disempowerment story in terms of irrationality. One could also think of wishful-thought attributions in the normatively thicker terms of epistemic irresponsibility. On this reading, Joe is representing Nkanyezi as epistemically irresponsible. But it doesn't much matter on which account of wishful thought we settle. On either reading, the explanation diminishes Nkanyezi's epistemic agency: it represents him as epistemically defective, as lacking the kind of rational authority over his belief that he in fact has.

5.3 The Reality of X-ist* Belief Explanation

One might worry at this point that the kind of phenomenon I have isolated here is so narrowly circumscribed, that even if it is occasionally instantiated, it is too rare to make it a worthwhile starting point for theorising the relationship between epistemic injustice and agency.

Unhappily, though, such explanations are all too widespread. Let me just mention three obvious areas. One is in everyday X-ist discourse as above. Thus, to recall my opening example, it is common for us to say that someone believes that we wronged her because she (like all girls) is overly sensitive. Such explanations are especially prevalent when the 'we' are privileged and the 'she' is not; and tend to represent the unprivileged group as for ever on the lookout for X-ist slight.

Another common kind of X-ist* explanation is ideological explanation. One often hears statements such as 'You only believe this because you are a white liberal', or 'She wouldn't believe this unless she were the out-and-out capitalist she is'. Such purely causal explanations will be sometimes correct. But when mistaken and driven by identity prejudice, they will often be instances of X-ist* belief explanation.

Finally, any incorrect 'interest explanation' – an explanation of a belief or theory in terms of interests rather than epistemic grounds – will follow the X-ist* pattern. For instance, sociologists of knowledge tend to explain the adoption of a certain scientific theory by appeal to the financial and social interests of the scientists involved. Such explanations tend to make the adoption of the theory sound irrational, undermining its scientific probity. My current proposal explains why: they give a causal explanation in place of a rationalising one. When the right explanation is the rationalising one, this constitutes a form of injustice.

6 The Place of Explanatory Injustice

I have argued here that the language of reasons illuminates the way in which our epistemic agency is wronged in explanatory injustice, the one involved in X-ist* explanation. A natural question to ask at this point is how explanatory injustice relates to the two more standard types of epistemic injustice in the literature, with which I opened.

I am not sure that there is any straightforward connection to hermeneutical injustice. This sort of injustice, recall, concerns depriving another of an epistemic resource that would enable her to make sense of her experiences either to herself or to others. Since X-ist* explanation doesn't confer or withhold this kind of resource, it seems unrelated to hermeneutical injustice.²⁶

²⁶ It should be noted, though, that explanatory injustice is probably one of the drivers of white ignorance (Mills 2007) and other forms of wilful hermeneutical ignorance (Pohlhaus 2012).

Some forms of testimonial injustice, however, could arguably be analysed in terms of X-ist* explanation. One obvious example is the *Talented Mr Ripley* case Miranda Fricker discusses. Briefly: Marge believes that Ripley is involved in Dickie's disappearance, based on witnessing Ripley's suspicious behaviour and being convinced that Dickie wouldn't commit suicide (the official hypothesis). Greenleaf explains her belief as due to 'a sweetheart's wishful thinking' (Fricker 2007: 87), silencing her with the words 'Marge, there's female intuition, and then there are facts' (cited by Fricker 2007: 9).²⁷ We can see this case as one in which a woman's testimony is disregarded due to an X-ist* explanation of her belief: Marge's belief is based on motivating reasons (evidence and a sound inference), but is explained instead in terms of a pure cause (wishful thought). Perhaps many kinds of testimonial injustice which involve doubting people's competence can be analysed in this way.²⁸ If so, the important thing to keep in mind is that explanatory injustice would be the genus. (Test: testimony was not doing the work in any of the examples until now.)

Thinking about some kinds of testimonial injustice as a species of explanatory injustice would have at least two advantages. First, it would allow us to see what is *culpable* about testimonial kinds of injustice. Fricker herself argues that the Greenleaf example involves *non-culpable* injustice, given the historical backdrop of the 1950's (2007: 100–1). But some philosophers have questioned this diagnosis, urging that injustice entails wronging someone (something Fricker herself concedes); but wronging entails culpability (e.g., Riggs 2012). If you share this sentiment, as I do, the above account gives you a straightforward way of pinning down what is culpable about this form of injustice: it is wrong to disempower another, to represent her as lacking her rightful epistemic authority, through ascribing a reason-why in place of the real motivating reason.

The second advantage to thinking about testimonial injustice as a kind of explanatory injustice, and hence, in terms of reasons, is that we can then see what is distinctively epistemic about testimonial injustice. My feeling is that existing accounts fall somewhat short in this respect. The thinking seems to be that what makes the injustice epistemic is that the victim is deprived of (non-incidentally) epistemic goods, such as credibility, and is thereby wronged *qua* knower (Fricker 2007: 1). But I don't think this is quite right. If you are a professional runner and I steal your only pair of trainers, I am unjustly depriving you of a sporting good and wronging you *qua* sportsman, but it doesn't follow that I have inflicted a sports injustice on you. Similarly, the move from the fact that I am depriving someone of an epistemic good to the claim that I have wronged him epistemically seems a bit quick.

If you agree and think that epistemic injustice is, nonetheless, a distinctively epistemic wrong, the present proposal helps account for this intuition: what is distinctively epistemic about explanatory injustice is that it involves the person's belittlement through robbing her of distinctively epistemic reasons. Such reasons aren't mere goods – like credibility and trainers. They are the stuff of epistemic normativity, the stuff that circumscribes the space of epistemic agency, and determines what you ought to believe, whether your belief counts as knowledge, and so on.²⁹

²⁷ In order not to overcomplicate things, I am happy that 'silencing' here is heard in the non-technical sense. But, of course, what goes on is also technical silencing. See, e.g., Dotson (2011).

²⁸ This is to distinguish it from cases, like Tom Robinson's, where the testifier's *sincerity*, rather than competence, is doubted. For a persuasive argument that these should be treated as different kinds of injustice, see McGlynn (2019).

²⁹ Admittedly, things are not as simple as I am presenting them here. (Thanks to an anonymous reviewer.) In particular, I am assuming that reasons are more basic to normativity than goods. And this is, of course, a contentious issue, and certainly not one that I can settle here. Hopefully, the unconvinced will at least accept that my account has the first advantage I claim it does.

Thus, thinking of X-ist* explanation not only illuminates a distinctive form of epistemic injustice, but also allows us to fruitfully analyse already acknowledged forms as a subspecies of the larger distinctive kind.

7 Conclusion

I have argued here that using the language of reasons allows us to spotlight a particular form of epistemic injustice involved in some biased explanations of belief. These are explanations which cite reasons-why when, in fact, the belief is based on motivating reasons. This constitutes an epistemic injustice, I have argued, because of the disempowering effect of reason-why explanations. First, I argued that confabulation of action is a form of self-empowerment precisely because it cites a motivating reason when only a reason-why is available. The empowerment is due to the fact that citing a motivating reason represents the actor as an agent of her action. Next, I showed that a certain kind of biased explanation of action does precisely the inverse: it disempowers the person we are explaining by citing a reason-why for her action, thus denying her agency in her own action. Third, I showed that the belief case is parallel for both confabulation and biased explanation. This suggested that when we explain someone's belief by citing a mere reason-why, there is a clear sense in which we are epistemically disempowering the person so explained, depriving her of epistemic authority.

This is no more than a sketch, of course, meant to open up a space for theorising epistemic injustice in relation to agency. Instead of filling in this sketch, though, I wish to close by gesturing at two potential doubts about this project. Allaying them will show us what we gain from thinking about explanatory injustice.

First, do we really need the language of reasons to say all I have said here? Couldn't we just say, 'Treating someone as irrational is epistemically wronging her by depriving her of her rightful epistemic status'? Well, of course we could say that, but we can't *just* say it. 'Irrational' and 'rightful epistemic status' are not very useful labels, because people mean so many different things by them. Isolating a particular kind of irrationality and associated epistemic status through a particular power-dynamic of reasons-ascription will hopefully be seen as giving us some finer tools for thinking about epistemic injustice.

But (and this is the second doubt) what is so special about the particular kind of biased explanation on which I have focused here? I have suggested several times that this kind of explanation of *action* is a fundamental form of disempowerment because it robs the agent of her agency over her action. An agent's action is something done for a reason. When we deny her that reason, we deprive her of her authorship of the action. A similar thing can be said about *belief*. A creature who was consistently at the mercy of mere causes would not have beliefs, it would be more like a thermostat than an agent. Interpreting my belief as being held for no good reason disempowers me by construing me, no matter how briefly, as such a creature. Given that such a creature is incapable of doing and believing things for reasons, and given that I am not such a creature, a fundamental sort of injustice has been inflicted on me.

If these thoughts are on the right track, the way is hopefully clear for an account of explanatory injustice and how it wrongs us *qua* epistemic agents.

Acknowledgments Huge thanks to Miloud Belkoniene, Ward Jones, Martin Kusch, and Robin McKenna, for really helpful comments on earlier drafts. Many thanks for their feedback on related talks, to my audiences at *Between Belief and Ethics Workshop* (Cologne 2019), *Epistemic Injustice, Reasons and Agency Workshop*

(Johannesburg 2019), the *24th World Congress of Philosophy* (Beijing 2018), *Epistemic and Practical Rationality Workshop* (Fribourg 2018), and *Rhodes Philosophy Graduate Day* (Rhodes University 2017). Thanks to the British Academy Advanced Newton Fellowship NAFR1180082 for their financial support.

References

- Alvarez M (2010) *Kinds of reasons: an essay in the philosophy of action*. Oxford University Press, New York
- Anscombe GE (1976) *Intention* (2nd edition, repr.). Blackwell, Oxford
- Broome J (1997) Reasons and motivation. *Aristotelian Society, Suppl* 77:131–146
- Buckwalter W, Stich S (2014) Gender and philosophical intuition. In: Knobe J, Nichols S (eds) *Experimental philosophy*, 2nd edn. Oxford University Press, Oxford, pp 307–346
- Campbell R, Kumar V (2012) Moral reasoning on the ground. *Ethics* 122(2):273–312
- Dancy J (2000) *Practical reality*. Oxford University Press, Oxford
- Davidson D (2001) *Essays on actions and events*, 2nd edn. Clarendon, Oxford
- Dotson K (2011) Tracking epistemic violence, tracking practices of silencing. *Hypatia* 26(2):236–257
- Dotson K (2012) A cautionary tale: on limiting epistemic oppression. *Front J Women Stud* 1:24–47
- Festinger L (1956) *A theory of cognitive dissonance*. Stanford University Press, Stanford
- Fricker M (2007) *Epistemic injustice: power and the ethics of knowing*. Oxford University Press, Oxford
- Hieronymi P (2005) The wrong kind of reason. *J Philos* 102(9):437–457
- Jackson F (1982) Review of the nature of causation by miles brand. *J Symb Log* 47(2):470–473
- Mackie JL (1965) Causes and conditions. *Am Philos Q* 2(4):245–264
- McGlynn A (2019) Epistemic objectification as the primary harm of testimonial injustice. *Episteme*:1–17. <https://doi.org/10.1017/epi.2019.9>
- Mills C (2007) White ignorance. In: Sullivan S, Tuana N (eds) *Race and epistemologies of ignorance*. State University of New York Press, Albany
- Mitova V (2017) *Believable Evidence*. Cambridge University Press, Cambridge
- Neta R (2002) S knows that P. *Nous* 36(4):663–681
- Newell BR, Shanks DR (2014) Unconscious influences on decision making: A critical review. *Behav Brain Sci* 37 (1):1–19
- Nisbett RE, Wilson TD (1977) Telling more than we can know: verbal reports on mental processes. *Psychol Rev* 84:231–259
- Parfit D (2001) Rationality and reasons. In: Dan E et al (eds) *Exploring practical philosophy*. Ashgate, Burlington, pp 17–39
- Peet A (2017) Epistemic injustice in utterance interpretation. *Synthese* 194(9):3421–3443
- Pohlhaus G (2012) Relational knowing and epistemic injustice: toward a theory of *Willful Hermeneutical Ignorance*. *Hypatia* 27(4):715–735
- Reisner A (2009) The possibility of pragmatic reasons for belief and the wrong kind of reasons problem. *Philos Stud* 145(2):257–272
- Riggs W (2012) Culpability for epistemic injustice: deontic or Aretetic? *Soc Epistemol* 26(2):149–162
- Summers JS (2017) Post hoc ergo propter hoc: some benefits of rationalization. *Philosophical Explorations* 20(sup1):21–36. <https://doi.org/10.1080/13869795.2017.1287292>
- van Fraassen B (1980) *The scientific image*. Oxford University Press, Oxford
- Weinberg J, Nichols S, Stich S (2001) Normativity and epistemic intuitions. *Philos Top* 29:429–460
- Williams B (1973) Deciding to believe. In: Williams B (ed) *Problems of the self*. Cambridge University Press, Cambridge
- Woodward J (2014) Scientific Explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), URL = <http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ethical Theory & Moral Practice is a copyright of Springer, 2020. All Rights Reserved.