Explaining Injustice in Speech: Individualistic vs. Structural Explanation

Saray Ayala (saray@sfsu.edu)
Department of Philosophy, 1600 Holloway Avenue
San Francisco, CA 94132 USA

Nadya Vasilyeva (Vasilyeva@Berkeley.Edu)
Department of Psychology, 3210 Tolman Hall
Berkeley, CA 94720 USA

Abstract

Implicit bias has recently gained much attention in scholarly attempts to understand and explain different forms of social injustice by identifying causally relevant mental states in individual' minds. Here we question the explanatory power of implicit bias in a particular type of injustice, testimonial injustice, and more generally in what we call *speech injustice*. Testimonial injustice occurs when the audience deflates a speaker's credibility due to the speaker's perceived social identity (Fricker, 2007). We identify two drawbacks of a widely accepted explanation attributing testimonial injustice to prejudices (e.g. implicit bias) in the mind of the hearer, and argue that further understanding of this phenomenon can be gained from a structural explanation that appeals to discursive conventions and interlocutors' positions in the communicative exchange.

Keywords: structural explanation, testimonial injustice, performative force, implicit bias, discursive injustice, social norms, norm-conforming behavior.

1. Deficiencies of an individualistic explanation

Imagine an all-men conversation about how to properly distribute the weight in the wooden structure of a house. A woman intervenes and proposes a solution. One interlocutor responds: "this is not home décor, this is serious work". This is an example of testimonial injustice, a subcategory of epistemic injustice (Fricker 2007). Through epistemic injustice, an individual is wronged in their capacity as a knower. In cases of testimonial injustice, an individual is wronged specifically in their capacity as a giver of knowledge, as happens to the woman in the example above.[1] In line with a recent trend in several disciplines focused on implicit bias as the key for the explanation of different forms of social injustice, a widely accepted explanation of testimonial injustice refers to the (explicit or implicit) attitudes of the individuals

involved. Miranda Fricker appeals to prejudices in the hearer that "will tend surreptitiously to inflate or deflate the credibility afforded the speaker" (Ibid. p.17), and specifies that "[T]he main type of prejudice (…) that tracks people in this way is prejudice relating to social identity" (Ibid. 27). This *biased mind* explanation, tracing particular episodes or patterns of injustice to bias in individuals' minds against speaker's social identity, is favored by all the psychological and philosophical literature drawing upon empirical research on implicit bias. A tacit assumption in at least some accounts of social injustice that appeal to implicit bias is that in the absence of implicit bias, the injustice would be reduced, or would not even happen.

Our goal in this work is not to question the existence or significance of implicit bias, but its explanatory power.[2] Explanations of injustice invoking implicit biases in individuals' minds could appropriately account for the phenomenon only in a society where interactions among individuals are *not* governed by unjust conventions or norms. If, however, the behavior is governed by conventions that program for and are sufficient to produce unjust treatment regardless of idiosyncrasies of individual attitudes, then biased-mind explanations alone can't adequately account for the contrast between just and unjust societies.

In this paper we outline ways in which structural explanations identifying higher-level social constraints on behavior address an important gap in our understanding of the phenomenon of social injustice. We spell out a structural explanation of testimonial injustice appealing to factors such as the discursive conventions that operate in the communicative exchange, and the positions that speaker and hearer(s) occupy in that communicative framework.[3] We identify two drawbacks of the *biased mind* explanation. First, it does not specify what happens

---

[1] Not all cases of differential attribution of credibility are cases of testimonial injustice, e.g., granting less credibility to a child than to an adult in a conversation about indexicals or law maximizes the communicative value of the conversation without disempowering any of the interlocutors.

[2] Our goal and motivation are similar to Haslanger's (2015b). She also argues that explanations of social injustice in terms of implicit bias are limited, and defends structural explanations instead. While she addresses social injustice in general, we focus on a specific type of injustice and offer some details of what a structural explanation of this type of injustice would look like.

[3] José Medina (2013) expands on Fricker's account of testimonial injustice, and points out the need of a less individualistic approach to it. His account nevertheless differs from the one we offer here.

with the speaker's words in cases of testimonial injustice; second, it only partially situates the individual in social reality. We discuss each of these drawbacks in detail below, and defend a structural explanation of testimonial injustice and of a broader form of injustice that we call *speech injustice*, which we define below.

1.1. What happens to the words?

The *biased mind* explanation does not tell us what happens with the speaker's words when their credibility is deflated. When a woman tries to assert something but her audience does not grant her the appropriate credibility, what is it that she gets to do with her words? We claim that Rebecca Kukla's notion of *discursive injustice* provides an answer (Kukla, 2014).[4] Discursive injustice occurs when a speaker's speech act is not given the appropriate uptake, in a way that disempowers the speaker. The performative force of the intended speech act gets distorted and the action that the speaker intended to bring about fails, and a different action occurs,[5] usually one that enhances a disadvantage, creates it and/or contributes to its perpetuation.[6] This distortion cannot be attributed to a failure of the speaker to use an appropriate locution.

Let's consider an example of testimonial injustice and analyze it with the tools of discursive injustice: a non-native speaker of English is attending an academic talk in an English-speaking country. The non-native speaker makes a contribution to the discussion (without any major deviations from standard English grammar and pronunciation that would interfere with comprehensibility of the intervention) but nobody engages with it. Eventually a native speaker raises the same point and this time people engage in a lively discussion.

In the example, a non-native speaker tries to make a proper contribution to a conversation, but fails to do so. There is plenty of research showing that non-native speakers are given less credibility than native speakers (Brown, Giles & Thakerar, 1985; Giles, 1973). Their perceived identity as a non-native speaker deflates their credibility. How does this happen? In general, when we make a contribution to a conversation, we assume that we are proper participants in that conversation, members of the discursive game that is being played. Our audience, however, might not consider us as such. The notion of *entreaty* helps understand this. Kukla describes *entreaties* as speech acts that are spoken "in order to be granted status as a speaker with normative standing within a

discursive subspace" (Kukla, 2014, p. 9). *Entreaties* come from outside the discursive game and are masked or direct requests: "Can I play too?". They do not count as (proper) contributions. If an utterance intended as a proper contribution is instead taken as an entreaty, this prevents the audience from evaluating its content in the same way that a proper contribution would be, for it only counts as a *request* to be part of the conversation. This suggests a possible way in which the non-native's credibility might be undermined: their intervention is not given uptake as a proper contribution to the discussion, but is rather treated as an entreaty, and so the lack of audience engagement is better seen as a response not to its content, but to its (perceived) performative force. If not recognized as a proper contribution to the conversation, there is no fact of the matter about whether or not the non-native's comment is worth engaging. This analysis of testimonial injustice as a distortion of the performative force clearly shows that it wrongs the speaker not only as a giver of knowledge, but also as a doer. When the force of a speech act is distorted in the aforementioned sense, speakers are prevented to do with their words what they intended to. Thus this principle captures a broader kind of injustice, beyond the epistemic kind. Let's call it *speech injustice*. Like discursive injustice as defined by Kukla (2004), speech injustice happens when a speaker's social position trumps the performative force of their speech acts, and they are prevented from doing with their words what they intended to; instead, their words bring about a different action, usually one that enhances a disadvantage, creates it and/or contributes to its perpetuation. Examples of speech injustice include assertions that are given uptake as expressives, or commands that are taken as requests.[7] Speech injustice differs from the phenomenon Kukla identifies in one important aspect: it is governed by conventions. While Kukla is not sympathetic with the idea that the discursive incapacity she is describing is a systematic phenomenon governed by conventions,[8] we emphasize that speech injustice feeds off discursive conventions. Our defense of a structural explanation of speech injustice in Section 2 captures exactly this idea.

1.2. Social situatedness

The second drawback of the *biased mind* explanation is that it does not appropriately situate the individual in their social reality. We draw on distributed and extended cognition research in the philosophy of mind (Menary, 2010; Sutton, 2010; Gallaguer, 2013) and defend a full-bodied situatedness, according to which the agent is in constant interaction with what is outside of their skin, continuously outsourcing to external (material and abstract) resources, be they technological aids or social norms. Although the *biased mind* explanation capitalizes on implicit biases and prejudices acquired by an individual from the society, it nevertheless isolates the

---

[4] Kukla (2014) does not introduce discursive injustice as the mechanism of testimonial injustice, but as an independent phenomenon.

[5] An extreme case of discursive injustice is Rae Langton's *illocutionary silence*, in which the speaker does not get to perform any action (Langton, 1993).

[6] The notion of discursive injustice is based on an important (and widely accepted) assumption according to which audience's uptake determines what one gets to do with their words (see Austin, 1962).

[7] See Kukla (2014) for several examples.

[8] See Kukla (2004), especially p. 448 and footnote 7.

individual from the social reality during the real time of the episode of injustice. A complete account of hearers' situatedness in the social reality must consider not only their attitudes and beliefs, even if they are a result of their exposure to social factors, but also the social factors themselves that, at that moment and in that space, are guiding their interpretation of speaker's words. Appealing exclusively to what is inside the agents' mind presents them isolated from the social dynamics and the structural forces that shape their behavior in that moment in important ways.

Once we situate the agent in this full-bodied sense, we need to pay attention to how environmental factors are not only facilitating (by providing norms as guides) but also constraining agents' behavior. When taking part in a conversation or any other social practice, we are part of a social structure. We draw on Sally Haslanger's notion of social structures as networks of social relations that are themselves constituted by practices (Haslanger, 2015a). These practices locate us in different positions in the structure, that Haslanger calls *nodes*. From each node, a limited range of possibilities is available (i.e. possible actions to choose from). Someone occupying the node of mother/spouse in a heterosexual relationship in a community with no affordable childcare, and with a male spouse earning a significantly higher salary due to a gender wage gap, has at their disposition a very limited range of possible actions from which to choose when a baby arrives. When the child-bearer decides to quit their job, this decision does not come from an unconstrained decision-making process that straightforwardly reflects their individual's preferences.[9]

Conversations, qua structured social interactions, situate interlocutors within nodes that afford limited ranges of actions. The affordances of each position within a conversation (e.g., speaker, hearer) are adjusted in virtue of membership in intersecting social categories, more or less salient. For example, a homeless person has a very limited range of actions they can do with their words when addressing a passerby. In an interview to apply for a visa to enter the United States, the status of authority and epistemic privilege of the interviewer, and the burden of proof of a lack of malicious intent lying upon the interviewee together reduce the range of possible speech acts the latter can perform. This range shrinks or expands depending on the combination of different dimensions of social location the person embodies (race, gender, socioeconomic status, level of education, immigration status, country of origin, etc.). For example, a citizen of the United States who tells an immigrant "Speak English, you are in the United States" would be enacting discrimination. If the speaker is however a homeless citizen and the addressee a middle-class immigrant, the speech act has a different force.

The *biased mind* explanation leaves aside all these structural constrains. In doing so, it paints a very limited picture of the factors governing speech injustice. We next argue that some of those structural factors better explain this kind of injustice.

2. Structural explanation

We mentioned above that the social position of an interlocutor in the broader social structure in which the conversational exchange is situated constrains the range of possible things this person can do with their words. There might be several structural factors playing an explanatory role for each particular social position, or in Haslanger's terms, for each node, and there are therefore different focuses a structural explanation can take. Here we don't provide a general analysis of what a structural explanation should look like in order for it to be satisfactory.[10] Our much more modest goal is to point out one type of structural factor that plays an explanatory role in speech injustice. And that is the conventions governing discourse. Conventions can be difficult to individuate and define. For our purposes, it is enough to say that there are conventional norms about what counts as a speech act of a particular type in a given context,[11] together with conventional ways to invoke those norms (e.g. particular words and gestures). In general we easily recognize, by interpreting words and gestures in any given conversational context, what are the conventions that our interlocutor is invoking, and we let those guide our interpretation of (the force of) their speech acts. We humans are in general very sensitive to norms; we are pretty good, from an early age, at detecting them, conforming to them right away, and punishing others for not doing so (Tomasello 2009).

We claim that our ordinary discursive interactions are governed by conventions that systematically distort the speech capacities of people perceived as occupying certain social positions. The existence of these conventions makes speech injustice a feature of well-functioning, although unjust, discursive exchanges. That is, speech injustice is not the result of occasional misinterpretation of some people's speech acts, but a systematic phenomenon resulting from conventions. Importantly, it is not the result of unskilled listeners who, due to (implicit) biases against speakers' identity, do not appropriately apply the relevant discursive conventions to that speaker's intervention. It is rather the unfortunate

---

[9] See Haslanger (2015a) for a detailed analysis of this example, originally in Cudd (2006).

[10] See Haslanger (2015b) for an analysis of three ways in which structures can be explanatory in accounting for social injustice.

[11] Very roughly, conventions are regularities we observe, upon which we agree either explicitly or implicitly, and which we deploy either consciously or unconsciously, and through which we organize and coordinate our behavior. Contrary to Lewis' traditional account (Lewis, 1969), and like many others, we don't take conventions to require common knowledge nor to be advantageous on all occasions (see e.g. Gilbert, 2008).

result of perfectly skilled listeners who are appropriately applying the conventions operative in their communities.

If we are right, people follow unjust conventions and we don't do anything about it. Could there be any other reason for that, apart from moral laziness? Yes. People are often wrong about the conventions they follow. Let us introduce a distinction that will help unpack this point.

2.1. Operative vs. Manifest conventions

We are sometimes mistaken about the concepts we apply. We might take ourselves to be applying a concept (the *manifest* concept) that does not actually coincide with the concept we apply in practice (the *operative* concept, see Haslanger, 2006). For example,

> "Consider the term "parent". It is common, at least in the United States, to address primary school memos to "Parents", to hold a "Parent Night" or "Parent Breakfast" at certain points during the school year, to have "Parent– Teacher Conferences" to discuss student progress, and so on. However, in practice the term "parent" in these contexts is meant to include the primary caregivers of the student, whether they be biological parents, step-parents, legal guardians, grandparents, aunts, uncles, older siblings, informal substitute parents, etc." (Haslanger, 2006, p. 99).

Even if the manifest concept used is the notion of immediate progenitor, the operative concept here is the notion of primary caregiver. If one only focuses on the manifest concept, they could be wrong about the concept they think to be applying.

Similarly, our intuitions about which discursive conventions operate in our linguistic communities might be wrong. Sometimes the mistake is easy to notice. On other occasions, however, we might not realize the error. Let's consider a fictional example: there is a universe with two planets, Earth and Twin Earth, inhabited by people who have antennae. On both planets there is the same manifest convention C about refusal of a sexual approach, let's say "to reject a sexual approach a person has to perform the standard formula of refusal, which is saying, by assertion, presupposition or implicature, 'no'". As it happens, on Twin Earth there is a subpopulation, let's call them antenna-movers, whose "no", when intended to refuse a sexual approach, does not get the uptake corresponding to a refusal, unless they utter it while moving their antennae. If they do not move their antennae, their "no" is interpreted in some other way. Thus the operative convention for refusal differs across the planets; on Twin Earth it depends on the perceived antenna-moving status of the speaker. We could define convention C* as a function of convention C *plus* speakers' perceived identity. On Twin Earth, the manifest convention for refusal (C) is a generally applicable convention, independent of identity. The operative convention (C*) is, however, constrained by the perceived identity of the speaker. For an antenna-mover on Twin Earth to successfully refuse a sexual approach, the standard formula is not enough, for their speech act won't be recognized as an act of refusal. They don't get to deploy the manifest convention C. Importantly, the gap between the manifest and the operative convention is not transparent to inhabitants of Twin Earth. Similarly to this fictional example, it can be the case that our ordinary discursive interactions are governed by *operative identity-constrained conventions* systematically distorting some speakers' speech acts.

It is clear now how the phenomenon we identify as speech injustice is different from the one Kukla identifies. She discards the possibility that the distortion of speaker's performative force is the result of some rules or conventions governing the conversational exchange. Referring to the resulting discursive incapacity, she says "If these effects were sufficiently regular they would become stable conventions in their own right, which could be managed and deployed in the normal way, even if they were politically unfortunate" (Ibid., p. 448). She assumes that the existence of (operative) conventions warrantees interlocutor's awareness of them, and subsequent capacity to deploy them in a controlled manner. However, once we acknowledge the possibility of operative conventions that might escape interlocutors' awareness, the existence of these conventions does not guarantee that speakers' will be able to manage and control their deployment.

Now, how does the existence of identity-constrained conventions help make the case for a structural explanation of speech injustice? One could say that even if we grant the existence of these conventions, they only work through the minds of the individuals, and so an explanation needs to appeal after all to individuals' mental states. But as we explain below, we don't need to posit any specific mental state (in particular, implicit bias against speaker's perceived social identity) to explain the behavior of hearers who conform to (operative) conventions.

2.2. Norm-conforming behavior

When we observe a behavior, the status of the behavior in relation to norms determines the extent to which that behavior can serve as a cue to the agent's mental states. Prescriptive norms give an agent enough of a reason to behave in accordance to them, and so only when someone behaves in a way that violates a norm, people go on to attribute to the agent some (other) reason to act that way, such as a norm-conflicting mental state (Uttich & Lombrozo, 2010). That is, the existence of a norm that prescribes a particular behavior in a given context can be seen as enough of a reason to exhibit that behavior in that context (no extra reason is required). To borrow an example from Uttich & Lombrozo, an explanation of why a person is wearing a graduation gown at a beach on a hot day (violating the norm) will need to invoke that person's mental states as reasons for the behavior (perhaps this person likes the gown, or believes the gown makes them look smart, etc.). In contrast, to explain why a person is

wearing a graduation gown at a graduation ceremony (adhering to the norm) it is not necessary to postulate a corresponding mental state. An explanation of a norm-conforming behavior does not have to invoke any mental state about the content of the norm - other than a general disposition to conform to it.[12] In fact, Uttich and Lombrozo suggest that is rational to follow such norm-based approach to explanation, for it affords accurate generalization and prediction about related cases.

Knobe (2007) showed that after observing people encouraging kissing between gay partners or interracial sex, subjects attributed intentions to them more often than to those who encouraged kissing between partners of different sexes or sex between partners of the same race. Although the original interpretation of this result appealed to unconscious bias against homosexual kissing and interracial sex, a different explanation seems to be better in accounting for these and other results.[13] Richard Holton proposes that we don't need to postulate such prejudices in order to explain subjects' attributions of intentions to the gay-kissing and interracial-sex fans. Subjects' awareness of social norms condemning those behaviors is enough (Holton, 2010). If gay-kissing and interracial-sex are seen as violating accepted social norms, those who engage in these activities or show sympathy for them will be seen as necessarily having some reason to do it. Those who conform to the norms don't need any extra reason (i.e. any specific mental state) for their heterosexual kissing and same-race sex preferences. In line with Holton's interpretation of Knobe (2007)'s results, we claim that we do not need to postulate implicit bias in hearers' minds in order to explain speech injustice; the existence of (unjust) social norms governing speech interactions is enough to explain this phenomenon.

2.2. What do we gain from identifying structural factors?

We argued that a structural explanation overcomes at least one of the drawbacks we identified in the individualistic explanation, i.e. the lack of situatedness. Moreover, a structural explanation is more stable. There are subtleties in prejudices and implicit bias that surely vary from one individual to the other, which make an explanation of any single episode of speech injustice to be limited to that episode. On the contrary a structural explanation, even though constrained by the cultural context and by eventual variations in discursive conventions, is more stable across individual episodes and

individual agents.[14] This stability makes structural explanations better at providing an understanding of speech injustice as a systematic social phenomenon.

Finally, we emphasize the importance of acknowledging structural factors for practical reasons. The presence of implicit bias only makes a difference to the phenomenon of speech injustice when there are no operative conventions systematically distorting speech acts. Effective interventions against speech injustice will have to take into account both individuals' minds and the conventions that constrain individuals' behavior.

3. Concluding remarks and directions for future research

We identified two deficiencies of the individualistic, *biased mind* explanation according to which testimonial injustice results from bias in the hearer's mind. We proposed that the first deficiency can be addressed with the notion of discursive injustice that appeals to distortion of performative force. This move revealed a kind of injustice that goes beyond the epistemic kind; we called this broader kind speech injustice.

To address the second drawback of insufficient situatedness in the social reality we brought to the scene the structural explanation. This explanation accounts for speech injustice in terms of unjust conventions that constrain interlocutors' behavior and the positions interlocutors take at any given conversation. We favor structural explanation of speech injustice for its stability and practical relevance, but we emphasize that it is compatible with the existence of implicit bias in hearers' minds. An important direction of future research will be to consider how structural and individual-level factors - discursive conventions operating in communicative frameworks and individuals' attitudes - interact in episodes of speech injustice, each party making the most of the other. For example, identity-constrained discursive conventions might become more powerful through individuals' implicit or explicit identity-prejudices; and more interestingly, individuals' attitudes might come to existence, and later be reinforced, through individuals conforming to unjust conventions in their discursive practices.

---

[12] We have to distinguish between (merely) following a norm and *intentionally* following it (see Holton, 2010). The latter might involve adjusting our behavior as to fit the norm, which would probably require beliefs about the content of the norm.

[13] Similar results from many other studies puzzled researchers. See Holton (2010) on why an explanation appealing to norm-conforming behavior accounts better for those results.

[14] See Haslanger (2015a) for more on why a structural explanation is in general more stable than an individualistic explanation.

References

Austin, J. L. 1962. *How to do things with words*, Oxford: Oxford University Press.

Brown, B. L., Giles, H., & Thakerar, J. N. 1985. Speaker evaluations as a function of speech rate, accent and context. *Language and Communication*, 5:207–220.

Dretske, F. 1988. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Fricker, M. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gallaguer, S. 2013. The Socially Extended Mind. *Cognitive Systems Research* 25-26:4-12.

Gilbert, M. 2008. Social Convention Revisited. *Topoi* 27:5–16.

Giles, H. 1973. Communication effectiveness as a function of accented speech. *Speech Monographs* 40; 330–331.

Haslanger, S. 2015a. What is a (Social) Structural Explanation? *Philosophical Studies* 172.

Haslanger, S. 2015b. Social Structure, Narrative, and Explanation. *Canadian Journal of Philosophy* 45 (1):1-15.

Haslanger, S. 2006. Philosophical Analysis and Social Kinds: What Good are Our Intuitions? *Proceedings of the Aristotelian Society*, Sup. Vol. 80:89– 118.

Holton, R. 2010. Norms and the Knobe Effect. *Analysis* 70 (3):417-424.

Keller, E. F. 2010. *The Mirage of a Space Between Nature and Nurture*. Durham, NC: Durham University Press.

Knobe, J. 2007. Reason Explanation in Folk Psychology. *Midwest Studies in Philosophy* 31:90-107.

Kukla, R. 2014. Performative Force, Convention, and Discursive Injustice. *Hypatia* 29 (2)

Langton, R.1993. Speech Acts and Unspeakable Acts, *Philosophy & Public Affairs* 22: 293-330.

Lewis, D. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Medina, J. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistance Imaginations*. Oxford: Oxford University Press.

Menary, R. (ed.) 2010. *The Extended Mind*. Cambridge, MA: MIT Press.

Sutton, J. 2010. Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In Menary, R. (ed.), *The Extended Mind*, MA: MIT Press.

Tomasello, M. 2009. *Why We Cooperate*. Cambridge, MA: MIT Press.

Uttich, K. & Lombrozo, T. 2010. Norms inform mental state ascriptions: a rational explanation for the side-effect effect. *Cognition* 116:87-100.