



Addressing Irrelevant Influences with Epistemic Humility

Joseph Gurrola

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

February 2, 2019

Addressing Irrelevant Influences with Epistemic Humility

Abstract

Implicit bias is an irrelevant influence which is pervasive in rational deliberation as well as in doxastic deliberation, and the doxastic attitudes informed by implicit biases have potentially far-reaching negative consequences. However, unlike other irrelevant influences, implicit biases are less likely to be identified by epistemic agents as factors which affect the rationality of doxastic attitudes since their presence in epistemic agents is difficult to identify by asking epistemic agents to reflect on singular beliefs or doxastic attitudes. Rather, identifying implicit biases as irrelevant influences in doxastic deliberation often requires the evaluation of many beliefs or actions of an epistemic agent to determine whether a pattern of bias exists.

Consequently, reactive approaches to addressing irrelevant influences in belief formation which ask agents to reflect on singular doxastic attitudes may be ill-suited to address irrelevant influences like implicit bias. In this paper, I offer a proactive solution for mitigating the negative effects of irrelevant influences like implicit bias, which requires that epistemic agents embrace epistemic humility in forming beliefs which they have good reason to think could be affected by irrelevant influences. In conjunction with reactive approaches to addressing irrelevant influences, epistemic humility about the extent to which epistemic agents' doxastic attitudes are informed by implicit bias and similar irrelevant influences has the potential to form better and more reliable epistemic agents in the long term.

Introduction

In this paper, I advance the thesis that implicit biases are irrelevant influences which impact our beliefs in nontrivial ways, and that, consequently, we should embrace epistemic

humility when forming beliefs which we have good reason to think could be negatively affected by implicit biases. My argument builds on Katia Vavova's Good Independent Reason Principle (**GIRP**) and argues that implicit bias is an irrelevant influence so prevalent in our epistemic practices that we have good independent reason to be proactive in ensuring our beliefs remain unaffected by implicit bias.

The first section of this paper defines essential terms. I begin by defining epistemic humility, as well as the prescriptive and descriptive elements of the epistemic humility I endorse. For purposes of clarity, I also define what I mean by "irrelevant influence". In section two, I discuss evidence for the descriptive claim which gives us reasons to think that implicit bias plays an important role in determining our beliefs. In section three, I engage Vavova's **GIRP** principle and argue that we have good reason to embrace epistemic humility as a means of guarding our beliefs against irrelevant influences like implicit bias. In part four I consider an objection to my argument.

I. Definitions: Epistemic Humility and Irrelevant Influences

Epistemic Humility

Epistemic humility is humility about the accuracy, reliability, consistency, or rationality of one's epistemic practices. Epistemic humility is akin to epistemic self-doubt in that epistemic humility typically requires that the epistemic agent take a critical stance towards her epistemic practices, to determine whether those practices are as accurate, reliable, consistent, or rational as she takes them to be (Roush). Epistemic humility is useful to the epistemic agent because its reflective methodology improves the degree to which the agent forms beliefs in accordance with their epistemic standards and epistemic values. Embracing epistemic humility also has the upshot

of making salient whether the agent behaves in accordance with her avowed beliefs, thus making clearer the extent to which the agent is the person she claims to be.

The Descriptive Claim

The descriptive aspect of the epistemic humility I endorse states that we often fall short of forming our beliefs solely by considering evidentially relevant factors. Though many agree that forming beliefs based on the evidence is the best method for forming reliably true beliefs, evidence about the ways irrelevant influences negatively affect our beliefs in important domains shows that we often fall short of the goal of forming beliefs based solely on the evidence.

The Prescriptive Claim

The prescriptive aspect of the epistemic humility I endorse holds that since we have good reason to think that irrelevant influences negatively affect our beliefs in important domains, we ought to engage in proactive reflection about the ways that beliefs formed in these domains could be negatively affected by irrelevant influences.

Irrelevant Influences

My formulation of irrelevant influences borrows from Katia Vavova's account of irrelevant influences which holds that irrelevant influences are influences which "don't bear on the truth of what we believe" (Vavova 1). Vavova further defines irrelevant influences as those which inform or "influence my belief that *P*" without bearing on the truth of *P* (3). Irrelevant influences in this sense can be something as seemingly innocuous as one's level of education, or where one was raised as a small child since some beliefs we have are no doubt affected by the beliefs or dispositions developed at our academic institutions or by immersion in the culture of our hometowns. For example, I may be more likely to espouse liberal values because I was raised in a predominately democratic state, or as a result of my training in the humanities.

Irrelevant influences can also be more pernicious in nature, such as having one's rational agency compromised by use of a drug which despite lacking a distinctive phenomenological effect nonetheless compromises one's epistemic practices in such a way that renders one unable to relinquish false beliefs in the face of relevant evidence¹. Each of these examples qualifies as an irrelevant influence because there are some beliefs which may be partially influenced by these factors, but whose truth does not depend on these factors.

In addition to these examples of irrelevant influences, I submit that we can also include in the set of irrelevant influences certain wide-reaching quantifiable biases, such as gender bias, or racial bias. These influences do not consciously enter the deliberative space of epistemic agents, yet they serve to influence beliefs in such a way that conflicts with agents' evidential standards, as well as with agents' epistemic values. Biases can affect beliefs even when an epistemic agent believes she is only attending to evidence in forming her beliefs. Tamar Gendler argues that biases seem to manifest without a distinctive phenomenology, thus being especially influential in generating beliefs without necessarily being a salient element in the process of rational deliberation (191). In these cases, one deliberates and weighs one's evidence in accordance with one's evidential standards and produces a belief which has been swayed by an irrelevant influence without one's. These irrelevant influences also conflict with one's epistemic values, since they result in actions and beliefs which manifest biased attitudes and actions that rational agents would explicitly disavow. For example, most rational agents disavow antiquated platitudes about the incompetence of women, yet may unknowingly harbor the kind of gender bias which results in a pattern of beliefs which, overall, reflect implicit gender bias. Given that

¹ Vavova describes such an example, in which a chef adds a certain anti-depressant to faculty members' food, thus rendering them recalcitrant to evidence against their false beliefs. (13).

even-handedness and objectivity are epistemic values that we consider essential to the set of values espoused by rational epistemic agents, biases which result in beliefs and actions that suggest contradictory values would be evidence against one's claim to be the epistemic agent one takes oneself to be. These biases are the irrelevant influences most pertinent to the current discussion.

For the purposes of this paper, irrelevant influences are those influences which affect the formation of doxastic attitudes without bearing on the truth of the propositions in question. These evidentially irrelevant factors are not themselves evidence for or against the propositions in question, and they may also not be salient components of doxastic deliberation insofar as they may operate in the background in the fashion of heuristics. Because they operate in the background of deliberative processes, these evidentially irrelevant factors may not be easily detectable without the assistance of third parties or objective evidence. In section three, I argue that techniques like reflection on one's deliberative practices are partly constitutive of the epistemic humility required to address these irrelevant factors.

II. Evidence for the Descriptive Claim

In this section, I briefly summarize evidence for the descriptive claim by reviewing an implicit bias study conducted by Tamar Gendler. The discussion of bias as an irrelevant influence will seek to show why bias is non-trivial in its effects on doxastic attitudes due to the far-reaching effects of biased doxastic attitudes.

Gendler (2014) discusses an implicit bias study in which content-identical CV's with identical cover letters indicating the candidate desired a one-year lab manager position were distributed to science faculty at 150 research institutions (194). Results showed that respondents

were likely to rate male candidates as significantly more competent and significantly more hireable than female candidates (Gendler 194).

Gendler's analysis of the study is that factors which we do not take to be relevant when forming our beliefs nonetheless affect our beliefs in a negative fashion. Gendler points out that "the effect in question is not driven by conscious or unconscious distaste," and cites as evidence for this claim the fact that respondents "tended to rate the female applicant as more likeable than the male applicant." (195) Gendler argues that this kind of bias can be explained by the fact that factors which do not play a conscious role in deliberation, such as one's upbringing and the norms of the society in which one was raised, do not have "a distinctive phenomenology²" and therefore are not salient in deliberation (195).

If Gendler's analysis is correct, then irrelevant factors such as one's upbringing and the norms of the society in which one lives have the potential to influence epistemic agents in non-trivial ways without their knowledge. These influences are non-trivial because the effects of implicit bias are far-reaching, and do much to determine not only the beliefs we form about others but also the way that those beliefs inform the way we behave toward others. If I, in considering whether to employ one of two equally qualified applicants of different genders find myself inclined to believe that the male candidate is more qualified, I may think this inclination is benign. But as Gendler's study shows, I am not the only one who will have this inclination, and many individuals will suffer because of my irrational belief.

When we combine the volume of our interactions with others with what we know about the prevalence of gender bias, the far-reaching effects of an irrelevant influence like gender bias

² By "distinctive phenomenology" Gendler means "a sort of internal mental discord or conflict" typical of influences which oppose our rational beliefs (191).

become clear: I am likely to form far more inaccurate beliefs because of my implicit biases. Since these false beliefs have far-reaching consequences³, we ought to exercise epistemic humility when forming and maintaining beliefs in domains where implicit bias may be especially relevant.

III. The Prescriptive Claim

While we can often identify certain irrelevant influences which affect our beliefs, Vavova argues that not all cases of irrelevant influence affecting belief are cases for concern. Instead, Vavova proposes the *Good Independent Reason Principle (GIRP)*, which holds that “to the extent that [one] has good independent reason to think” one’s doxastic attitude with respect to *P* has been compromised, one must revise one’s confidence in one’s doxastic attitude with respect to *P* “accordingly—insofar as [one] can⁴” (12). Vavova contrasts this principle with the *No Independent Reason Principle (NIRP)*, which holds that one must revise one’s doxastic attitude if one lacks good independent reason to think that one is *not* mistaken in their doxastic attitude toward a proposition. This latter principle is problematic, however, since it requires that agents provide independent justification for the thought processes and background assumptions which resulted in their doxastic attitudes, but it is impossible for agents to provide reasons for their attitudes independent of the background assumptions and the original reasons given for the doxastic attitudes. Since any principle which requires this level of independent justification on the part of epistemic agents is not especially useful in distinguishing between cases of genuinely

³ In mentioning the far-reaching consequences of biases, I am thinking of explicit racism and sexism, as well as forms of micro-aggression.

⁴ Vavova expresses **GIRP** simply in terms of one’s relationship to *P*, rather than one’s doxastic attitude. I employ this more specific characterization for the sake of clarity. In my characterization, “doxastic attitude” includes cases of belief, suspension of belief, and cases of disbelief.

problematic irrelevant influences and innocuous ones (since it requires independent justification in all cases), a principle like GIRP is more beneficial for determining when agents need to worry about their doxastic attitudes being affected by irrelevant influences.

Vavova argues that by adopting GIRP, agents can maintain their general sense of authority over their doxastic attitudes, while also embracing the responsibility for reflection when they encounter evidence of error or of irrelevant influence. Importantly, reflection itself does not require the agent to alter her confidence in her doxastic attitude. The agent is only required to reduce confidence in her doxastic attitude if, upon reflection, she finds that she has good independent reason to think that the irrelevant influences in question have put her in a bad epistemic situation (Vavova 11).

While Vavova's principle provides a useful means of reflection when encountering evidence of error, it is not clear to what extent this principle is applicable to doxastic attitudes which, though negatively affected by irrelevant influences, are less likely to be identified as cases of error since they may be considered uncontroversial beliefs when evaluated on their own merits. For example, doxastic attitudes that are affected by implicit bias are typically not identified as such by the agent or by those with whom the agent generally interacts. This is what makes implicit bias so problematic—the agents manifesting these biases interact with others who are equally likely to manifest these biases, and equally likely to be oblivious to the patterns of doxastic attitudes caused by these biases over the long term. Only by taking the long view, and attending to the patterns of beliefs and actions over time, can agents gain evidence of error caused by implicit bias and initiate reflective processes. This is problematic because individual cases of epistemic irrationality due to bias are less likely to be identified without knowledge of the agent's doxastic attitudes and actions in the long term. Since it is often the case that no

singular belief affected by implicit bias can provide evidence of error on its own, GIRP may not be sufficient for addressing errors generated by irrelevant influences like implicit bias.

A related concern is that an agent may lack good independent reasons to think that her doxastic attitudes affected by bias are genuinely informed by bias. She may claim that, upon reflection, bias did not play a role in her reasoning, and that her belief was based solely on evidentially relevant factors. We can imagine the scientists in Gendler's study providing an analogous explanation of their actions if asked whether bias informed their doxastic attitudes.

If implicit biases are irrelevant influences we ought to worry about, and they are not the kinds of irrelevant influences which are easily detected in reflection about individual doxastic attitudes, then a reactive principle like GIRP may be limited in its ability to address irrationality generated by implicit biases. To address this concern, I submit that we adopt a principle of epistemic humility which acts in conjunction with GIRP to address those irrelevant influences that place agents in bad epistemic situations, but which are not necessarily detectable in reflection about singular beliefs. This principle holds that since agents have good independent reason to believe that their beliefs are often informed by implicit bias and that their bias is likely not detectable in any one action or belief, they are required to be proactive in guarding against implicit bias. In guarding against implicit bias, agents must engage in reflective thought about their epistemic practices when forming beliefs which they know may be negatively affected by bias. Agents may also employ other tactics in proactively accounting for error, such as enlisting those around them to act as "error detectors", and seeking out objective evidence about the domains in which their beliefs are most frequently affected by implicit bias.

In conjunction with GIRP, epistemic humility about the extent to which one's doxastic attitudes are informed by implicit bias and similar irrelevant influences which are only detectable

in the long term has the potential to form better and more reliable epistemic agents in the long term. If an agent espouses GIRP and commits herself to reflection in cases of error, and also espouses epistemic humility about implicit bias, then the agent has a better chance of being the epistemic agent she claims to be. Battling existing heuristics with consciously adopted heuristics like GIRP and epistemic humility provides a robust framework of reactive and proactive principles to guide an epistemic agent in forming the most reliable beliefs.

IV. Is Implicit Bias a Genuine Irrelevant Influence?

One may object to my inclusion of implicit bias into the set of irrelevant influences, on the grounds that psychological facts about us which negatively impact rationality are not themselves irrelevant influences since they are part of our rational makeup. Therefore, considering these kinds of bias or dispositions as irrelevant influences may make the category of irrelevant influences too broad.

While these biases and heuristics are part of our rational makeup, they are nonetheless things which we can analyze and improve if necessary. That is, we can address and modulate these heuristics without fundamentally changing our rational makeup. Additionally, social biases are special in that they are heuristics we feel compelled to address, whereas other heuristics, like hyperactive agency detection (HAD), are not heuristics we necessarily feel compelled to address or improve, despite their propensity to generate false beliefs. In addition to being fitness-enhancing, HAD is not in any important sense at odds with the way that we think about ourselves as epistemic agents. Epistemic rationality is compatible with some more primitive elements of human psychology which don't seem to impugn the claim that we are rational creatures. Unlike HAD, irrelevant influences like implicit bias do impugn our conception of ourselves as rational epistemic agents who form beliefs based off our evidential standards. Thus, it seems like some

forms of built-in irrationality can bear more directly on our claim to be rational epistemic agents than others.

V. Conclusion

In this paper, I argued that implicit biases are especially problematic irrelevant influences for two reasons. First, they have far-reaching effects through their influence on our beliefs. Second, they are not irrelevant influences which are easily detectable, since the beliefs they affect are not on their own likely to produce evidence of error. Consequently, I proposed adopting epistemic humility in the form of reflection as an epistemic heuristic to mitigate the negative effects of implicit bias as an irrelevant influence. Adopting epistemic humility as a proactive means of battling implicit bias has the upshot of creating a more reflective epistemic agent, and provides a robust means for addressing an especially harmful irrelevant influence that may not be addressable by heuristics which only activate after epistemic agents are presented with evidence of error.

Works Cited

Gendler, Tamar Szabó. "I—The Third Horse: On Unendorsed Association and Human Behaviour." *Aristotelian Society Supplementary Volume*, vol. 88, no. 1, June 2014, pp. 185–218, doi:10.1111/j.1467-8349.2014.00239.x.

Roush, Sherrilyn, "Epistemic Self-Doubt", *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/win2017/entries/epistemic-self-doubt/>](https://plato.stanford.edu/archives/win2017/entries/epistemic-self-doubt/).

Vavova, Katia. "Irrelevant Influences." *Philosophy and Phenomenological Research*, vol. 96, no. 1, Jan. 2018, pp. 134–52, doi:10.1111/phpr.12297.