

The Epistemology of Implicit Biases

Carlotta Pavese

11.11.14

Outline

Introduction

The source of implicit biases

From implicit biases to skepticism

Aliefs

Outline

Introduction

The source of implicit biases

From implicit biases to skepticism

Aliefs

What is an implicit bias?

- ▶ You must have taken some of the tests at:
<https://implicit.harvard.edu/implicit/>.

What is an implicit bias?

- ▶ You must have taken some of the tests at:
<https://implicit.harvard.edu/implicit/>.
- ▶ On the basis of those tests, can you make an educated guess as to what an implicit bias is?

What is an implicit bias?

There is a vast and still-growing literature on implicit bias, which I can only be dipping into here. Very broadly speaking, these are largely unconscious tendencies to automatically associate concepts with one another.³ Put like this, they don't sound very interesting or worrying. But the ones on which attention by philosophers has focused are both very interesting and very worrying. These are unconscious, automatic tendencies to associate certain traits with members of particular social groups, in ways that lead to some very disturbing errors: we tend to judge members of stigmatized groups more negatively, in (Saul 2013, p. 144).

Examples

Curriculum vitae

CV studies take a common, and beautifully simple form. The experimenters ask subjects to rate what is in fact the same CV, varying whatever trait they want to study by (usually) varying the name at the top of it. When they do this, they find that the same CV is considered much better when it has a typically white rather than typically black name, a typically Swedish rather than typically Arab name, a typically male rather than typically female name, and so on. The right name makes the reader rate one as more likely to be interviewed, more likely to be hired, likely to be offered more money, and a better prospect for mentoring. These judgments are very clearly being affected by something that *should* be irrelevant — the social category of the person whose CV is being read. Moreover, the person making these mistaken judgments is surely unaware of the role that social category is playing in the formation of their views of the candidates.

(Saul 2013, p. 144).

Examples

Prestige bias

In a now-classic study, psychologists Peters and Ceci (1982) sent previously published papers to the top psychology journals that had published them, but with false names and non-prestigious affiliations. Only 8% detected that the papers had already been submitted, and 89% were rejected, citing serious methodological errors (and not the one they should have cited — plagiarism). This makes it clear that institutional affiliation has a dramatic effect on the judgments made by reviewers (either positively, negatively, or both). These are experts in their field, making judgments about their area of expertise — psychological methodology — and yet they are making dramatically different judgments depending on the social group to which authors belong (member of prestigious VS non-prestigious psychology department).

(Saul 2013, p. 145).

Examples

Perception

Studies of so-called ‘shooter bias’ show us that implicit bias can even influence perception. In these studies, it has been shown that the very same ambiguous object is far more likely to be perceived as a gun when held by a young black man and something innocent (like a phone) when held like by a young white man.⁵ (The same effect has been shown with men who appear Muslim versus men who appear non-Muslim (Unkelbach et al. 2008). In some of these experiments, the subjects’ task is to shoot in a video game if and only if they see an image of a person carrying a gun. Subjects’ ‘shooting’ is just as

(Saul 2013, p. 146).

How are implicit biases different from (explicit) beliefs?

- ▶ We are not aware of them.

How are implicit biases different from (explicit) beliefs?

- ▶ We are not aware of them.
- ▶ We may even disavow them if asked.

How are implicit biases different from (explicit) beliefs?

- ▶ We are not aware of them.
- ▶ We may even disavow them if asked.
- ▶ It is not clear that they are rationally revisable in the way beliefs are.

How are implicit biases different from (explicit) beliefs?

- ▶ We are not aware of them.
- ▶ We may even disavow them if asked.
- ▶ It is not clear that they are rationally revisable in the way beliefs are.
- ▶ Anything else?

Moral and political consequences of implicit biases

- ▶ We are very likely to make inaccurate judgments about who is the best candidate for a job, if some of the top candidates are known to be from stigmatized groups.

Moral and political consequences of implicit biases

- ▶ We are very likely to make inaccurate judgments about who is the best candidate for a job, if some of the top candidates are known to be from stigmatized groups.
- ▶ We are very likely to mark inaccurately, if social group membership is known to us and the group we are marking is not socially homogeneous.

Moral and political consequences of implicit biases

- ▶ We are very likely to make inaccurate judgments about who is the best candidate for a job, if some of the top candidates are known to be from stigmatized groups.
- ▶ We are very likely to mark inaccurately, if social group membership is known to us and the group we are marking is not socially homogeneous.
- ▶ We are very likely to make inaccurate judgments about which papers deserve to be published, if social group membership is known to us.

Outline

Introduction

The source of implicit biases

From implicit biases to skepticism

Aliefs

Why do we form implicit biases

THE COMPLEXITY OF THE WORLD

Categorization and stereotyping are tools used by finite minds to operate effectively in an overwhelmingly complex environment. Cognizing entities in ways that involve classifying them as members of larger collections solves two problems simultaneously. On the one hand, it allows us to navigate a world whose complexity exceeds our cognitive capacities: the world is complicated, our cognitive resources are limited, and classifying objects into groups allows us to proceed effectively in an environment teeming with overwhelming detail. On the other hand, it allows us to navigate a world where in any given encounter, an object presents us with only a few of its potentially relevant features. Classifying objects into groups allows us to readily extrapolate properties that that object is likely to share with other members of its kind¹⁴. So categorization is not an optional way of making sense of the world: it is our means of dealing with the problems of global complexity and experience poverty.

Given this, it's perhaps unsurprising that categorization is something that comes on line

(Gendler, p. 6).

The sources of implicit biases

- ▶ One source of implicit biases is that because categorization is driven by a goal of simplicity, there is a tendency to view individuals within a category in ways that emphasize their similarities, and view individuals between categories in ways that emphasize their differences; that is, there is a tendency towards **intracategory assimilation**, and **intercategory contrast**.

The sources of implicit biases

- ▶ There is a process of **confirmation bias**:
Stereotype-congruent information is attended to and encoded; stereotype-incongruent information is ignored and unassimilated. Both the search for and interpretation of information tends to be done in ways that favor hypotheses that the subject already holds.

Aversive racism

- ▶ An example of stereotypical association: *aversive racism*.

Aversive racism

- ▶ An example of stereotypical association: *aversive racism*.
- ▶ The members of a minority group are associated with a stereotype, that includes a set of traits.

Aversive racism

- ▶ An example of stereotypical association: *aversive racism*.
- ▶ The members of a minority group are associated with a stereotype, that includes a set of traits.
- ▶ These characteristics were identified as constituting the cultural stereotype even by those who rejected their accuracy.

Aversive racism

- ▶ An example of stereotypical association: *aversive racism*.
- ▶ The members of a minority group are associated with a stereotype, that includes a set of traits.
- ▶ These characteristics were identified as constituting the cultural stereotype even by those who rejected their accuracy.
- ▶ Knowledge of the stereotype does not need to correlate in any way with measures of explicit prejudice.

Aversive racism

- ▶ “aversive racism”—characteristic of those who “sympathize with the victims of past injustice, support the principle of racial equality, and regard themselves as non-prejudiced”

Aversive racism

- ▶ “aversive racism”—characteristic of those who “sympathize with the victims of past injustice, support the principle of racial equality, and regard themselves as non-prejudiced”
- ▶ but who, because of their explicit or implicit awareness of the negative traits stereotypically associated with members of the dominated racial group, exhibit behaviors indicative of negative feelings and beliefs about blacks, which may be unconscious”.

Example of Aversive racism

- ▶ Even among those who are explicitly and sincerely committed to anti-racism, the legacy of having lived in a society structured by hierarchical and hostile racial divisions retains its imprint.

Example of Aversive racism

- ▶ Even among those who are explicitly and sincerely committed to anti-racism, the legacy of having lived in a society structured by hierarchical and hostile racial divisions retains its imprint.
- ▶ Likewise, participants playing a video game are quicker at deciding to shoot an unarmed black target than an unarmed white target, even when both targets are armed at equal rates in the context of the game (Correll et al., 2002).

Example of Aversive racism

- ▶ State legislators are less likely to respond to requests for help with voter registration when the requests come from individuals with stereotypically black names (Butler Broockman, 2011).

Example of Aversive racism

- ▶ State legislators are less likely to respond to requests for help with voter registration when the requests come from individuals with stereotypically black names (Butler Broockman, 2011).
- ▶ Black cab drivers receive lower tips than white cab drivers (Ayres et al., 2004).

Example of Aversive racism

Indeed, even those who devote their lives to counteracting such stereotypical associations are not immune from them. Consider the following quotation:

There is nothing more painful to me at this stage in my life than to walk down the street and hear footsteps and start thinking about robbery. Then look around and see somebody White and feel relieved.

Its utterer is none other than Jesse Jackson ("Jesse Jackson" 2011).

Outline

Introduction

The source of implicit biases

From implicit biases to skepticism

Aliefs

Jennifer Saul

Cast of Characters



- ▶ Philosopher of language at University of Sheffield (UK)
- ▶ PhD in philosophy at Princeton University.
- ▶ Her most recent book is *Lying, Misleading and What is Said: An Exploration in Philosophy of Language and in Ethics* (Oxford University Press 2012).

Tamara Gendler

Cast of Characters



- ▶ Professor of Philosophy at Yale University.
- ▶ PhD in philosophy at Harvard University.
- ▶ Author of many books, among which *Intuition, Imagination and Philosophical Methodology*, 2010; *The Elements of Philosophy: Readings from Past and Present*; *Conceivability and Possibility*; *Thought Experiment: On the Powers and Limits of Imaginary Cases*.

from Implicit biases to skepticism

- ▶ Implicit biases have clear epistemic costs.

from Implicit biases to skepticism

- ▶ Implicit biases have clear epistemic costs.
- ▶ On their basis, we form beliefs that are likely to be wrong.

from Implicit biases to skepticism

now, some less obvious epistemological aspects of the situation, again focussing on philosophy. When we misjudge a paper's quality, we're making a mistake about the quality of an argument.⁷ Moreover, our evaluation of that argument is being influenced by factors totally irrelevant to its quality: it's being influenced by our knowledge of the social group of its author. Worse yet, this influence operates below the level of consciousness — it's unavailable to inspection and rational evaluation. This means we may be accepting arguments we should not accept and rejecting arguments we should not reject. Many of our philosophical beliefs — those beliefs we take to have been arrived at through the most careful exercise of reason — are likely to be wrong.⁸

from Implicit biases to skepticism

- ▶ Jennifer Saul thinks that the widespread presence of implicit bias gives rise to a **distinctive kind of skepticism**, different from the traditional one.

from Implicit biases to skepticism

- ▶ Jennifer Saul thinks that the widespread presence of implicit bias gives rise to a **distinctive kind of skepticism**, different from the traditional one.
- ▶ It is different partly because of its clear practical relevance.

from Implicit biases to skepticism

entific or argumentative merit.

But why should that unsettle us? We know already that most of what is currently accepted as science is likely to be proven false within centuries, and possibly decades. But notice: my claim is not that we're likely to be accepting some falsehoods, or even a lot of falsehoods. That's not unsettling. My claim is that we're likely to be *making errors*. Moreover, we're likely to be making errors of a very specific sort. It's *not* that we're likely to get some really difficult technical bits wrong, or that we're likely to get things wrong if we're really exhausted, or drunk. It's that we're likely to let the social identity of the person making an argument affect our evaluation of that argument. It is part of our self-understanding as rational enquirers that we will make certain sorts of mistakes. But not this sort of mistake. These mistakes are ones in which something that we actively think *should not* affect us does.

from Implicit biases to skepticism

- ▶ Even though implicit biases are not themselves epistemically assessable, they give rise to beliefs that are.

from Implicit biases to skepticism

- ▶ Even though implicit biases are not themselves epistemically assessable, they give rise to beliefs that are.
- ▶ Beliefs based on implicit biases are very likely to be wrong.

from Implicit biases to skepticism

- ▶ Even though implicit biases are not themselves epistemically assessable, they give rise to beliefs that are.
- ▶ Beliefs based on implicit biases are very likely to be wrong.
- ▶ That means that in a variety of cases, the implicit biases we do have are likely to lead us astray from the truth.

Gendler on The costs of implicit biases

- ▶ The costs are ones that arise simply from having encoded the stereotypes, whether or not the subject endorses them.

Gendler on The costs of implicit biases

- ▶ The costs are ones that arise simply from having encoded the stereotypes, whether or not the subject endorses them.
- ▶ Gendler argues that the costs are diverse.

First kind of cost

- ▶ It affects your ability to encode individuating information about faces of persons that you apprehend as belonging to a different racial group—even if you explicitly avow racial equality.

First kind of cost

- ▶ It affects your ability to encode individuating information about faces of persons that you apprehend as belonging to a different racial group—even if you explicitly avow racial equality.
- ▶ Across hundreds of studies in dozens of cultures, psychologists have shown a tendency for participants to exhibit **superior recognitional capacities for own-race** as compared to other-race faces (cf. Meissner Brigham 2001).

First kind of cost

- ▶ **Cross-race recognition deficit** is the phenomenon whereby faces of the out-group are recognized (misses) or remembered (false alarms) less well than faces of the in-group.

First kind of cost

- ▶ **Cross-race recognition deficit** is the phenomenon whereby faces of the out-group are recognized (misses) or remembered (false alarms) less well than faces of the in-group.
- ▶ According to the **asymmetric feature selection** hypothesis, at least part of the explanation is the following: when participants encounter other-race faces, one of the visual features they typically encode is information about race, whereas no such information is typically encoded for same-race faces. This leaves her with more cognitive space to encode an additional fact about the same-race face—say, “eyebrows there.”

First kind of cost

- ▶ So, for example, when a white subject sees a novel black face, in addition to coding information like “eyes here”, “nose there”, “ears there,” she also uses some of her limited cognitive resources to encode “black;?”

First kind of cost

- ▶ So, for example, when a white subject sees a novel black face, in addition to coding information like “eyes here”, “nose there”, “ears there,” she also uses some of her limited cognitive resources to encode “black;?”
- ▶ whereas, typically, when she sees a novel white face, the category of race is not encoded as such.

First kind of cost

- ▶ Since other-race faces are processed as racially marked, cognitive resources that would otherwise be available for encoding specific information about the face are deployed to encode coarse-grained information about category-membership;

First kind of cost

- ▶ Since other-race faces are processed as racially marked, cognitive resources that would otherwise be available for encoding specific information about the face are deployed to encode coarse-grained information about category-membership;
- ▶ Since same-race faces are processed as racially unmarked, the subject's limited cognitive resources can be deployed to encode more fine-grained information.

A second epistemic cost: Stereotype threat

- ▶ The presence of implicit biases impairs cognitive performance when a negative stereotype about a group with which you self-identify is brought to salience—even if you explicitly believe those negative associations to be false.

A second epistemic cost: Stereotype threat

- ▶ The presence of implicit biases impairs cognitive performance when a negative stereotype about a group with which you self-identify is brought to salience—even if you explicitly believe those negative associations to be false.
- ▶ **Stereotype threat** is a well-documented phenomenon whereby activating an individual's thoughts about her membership in a group that is associated with impaired performance in a particular domain increases her tendency to perform in a stereotype-confirming manner.

A second epistemic cost: Stereotype threat

In one of the most striking demonstrations of the phenomenon, young girls of Asian- American descent who ranged in age from kindergarten to 8th grade were given tasks that rendered salient either their female identity, their Asian identity, or neither identity (control). Subsequently, they were given a series of items from a grade-appropriate standardized math test. Girls from lower-elementary and middle school grades showed a striking pattern of results: those whose Asian identity had been emphasized showed an improvement in scores when compared with controls, whereas those whose female identity had been emphasized showed a decrement.

A second epistemic cost: Stereotype threat

- ▶ Stereotype threat appears to interfere with knowledge in at least two ways.

A second epistemic cost: Stereotype threat

- ▶ Stereotype threat appears to interfere with knowledge in at least two ways.
- ▶ Participants may temporarily lose access to the contents of certain of their true beliefs.

A second epistemic cost: Stereotype threat

- ▶ Stereotype threat appears to interfere with knowledge in at least two ways.
- ▶ Participants may temporarily lose access to the contents of certain of their true beliefs.
- ▶ And participants may temporarily lose confidence in their true beliefs.

Issues for discussion

- ▶ If they are implicit, and may be unaware of them, in what sense are we (morally and epistemically) responsible for them?

Issues for discussion

- ▶ If they are implicit, and may be unaware of them, in what sense are we (morally and epistemically) responsible for them?
- ▶ Possible replies: individually, we may be unaware of them, but we are aware that they exist because there is plenty of psychological evidence for their existence.

Issues for discussion

- ▶ If they are implicit, and may be unaware of them, in what sense are we (morally and epistemically) responsible for them?
- ▶ Possible replies: individually, we may be unaware of them, but we are aware that they exist because there is plenty of psychological evidence for their existence.
- ▶ Although we may have no control on them as individual, there may be ways in which we may control their developments by changing certain features of society.

Overcoming implicit biases

- ▶ Can implicit biases, and the skepticism they give rise to, be overcome?

Overcoming implicit biases

- ▶ Can implicit biases, and the skepticism they give rise to, be overcome?
- ▶ How could we overcome implicit biases?

Overcoming implicit biases

- ▶ The problem is: if the processes by which we acquire implicit biases are not under our control, it is not clear that more reflectiveness on our beliefs and behavior will actually help.

Overcoming implicit biases

- ▶ The problem is: if the processes by which we acquire implicit biases are not under our control, it is not clear that more reflectiveness on our beliefs and behavior will actually help.
- ▶ After all, if we are wired to acquire implicit biases from the world surrounding us, then it looks like we need to alter the world itself.

Overcoming implicit biases

inevitably limited.

To fully combat the influence of implicit biases, what we really need to do is to re-shape our social world. The stereotypes underlying implicit biases can only fully be broken down by creating more integrated neighborhoods and workplaces; by having women, people of colour and disabled people in positions of power; by having men in nurturing roles; and so on. The only way to be fully freed from

Outline

Introduction

The source of implicit biases

From implicit biases to skepticism

Aliefs

Gendler's Aliefs = implicit biases?

useful to say a few words about the concept²¹. As I noted above, to have an *alief* is to have an innate or habitual propensity to respond to an apparent stimulus in a particular way. In paradigm cases and on strict usage, this response involves an automatized representational-affective-behavioral triad. I'll run through a specific case with this strict characterization, and then say something about how the notion generalizes²².

When I stand on a transparent glass walkway projecting over the Grand Canyon, this

Example I of alief

Below the floor's five layers of glass (protected from scratches by the booties) can be seen the cracked, sharp-edged rock face of the canyon's rim and a drop of thousands of feet to the chasm below. The promise is the dizzying thrill of vertigo. And indeed, last week some visitors to this steel-supported walkway anchored in rock felt precisely that. One woman, her left hand desperately grasping the 60- inch-high glass sides and the other clutching the arm of a patient security guard, didn't dare move toward the transparent center of the walkway.

Comments

How should we describe the cognitive state of those who manage to stride to the Skywalk's center? Surely they believe that the walkway will hold: no one would willingly step onto a mile-high platform if they had even a scintilla of doubt concerning its stability. But alongside that belief there is something else going on. Although the venturesome souls wholeheartedly believe that the walkway is completely safe, they also alieve something very different. The alief has roughly the following content: "Really high up, long long way down. Not a safe place to be! Get off!!"

Example II of alief

Charles is watching a horror movie about a terrible green slime. He cringes in his seat as the slime oozes slowly but relentlessly over the earth destroying everything in its path. Soon a greasy head emerges from the undulating mass, and two beady eyes roll around, finally fixing on the camera. The slime, picking up speed, oozes on a new course straight towards the viewers. Charles emits a shriek and clutches desperately at his chair.

Comments

How should we describe Charles' cognitive state? Surely he does not believe that that he is in physical peril; as Walton writes "Charles knows perfectly well that the slime is not real and that he is in no danger." But alongside that belief there is something else going on. Although Charles believes that he is sitting safely in a chair in a theater in front of a movie screen, he also alieves something very different. The alief has roughly the following content: "Dangerous two-eyed creature heading towards me! H-e-l-p ... ! Activate fight or flight adrenaline now!"

Characterization of aliefs

alief is associative, action-generating, affect-laden, arational, automatic, agnostic with respect to its content, shared with animals, and developmentally and conceptually antecedent to other cognitive attitudes.

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?
- ▶ The content of aliefs seems to serve some evolutionary purpose—protect us from danger, for example.

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?
- ▶ The content of aliefs seems to serve some evolutionary purpose—protect us from danger, for example.
- ▶ No similar justification there can be for the content of *implicit biases*.

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?
- ▶ The content of aliefs seems to serve some evolutionary purpose—protect us from danger, for example.
- ▶ No similar justification there can be for the content of *implicit biases*.
- ▶ It is interesting and puzzling that Gendler mixes together aliefs and implicit biases under the same heading. Do not you think?

Aliefs/implicit biases

- ▶ In an important sense, we do not seem to be responsible for being afraid of height.

Aliefs/implicit biases

- ▶ In an important sense, we do not seem to be responsible for being afraid of height.
- ▶ No we should try to revise those aliefs, because it is helpful to have those kind of implicit fears.

Aliefs/implicit biases

- ▶ In an important sense, we do not seem to be responsible for being afraid of height.
- ▶ No we should try to revise those aliefs, because it is helpful to have those kind of implicit fears.
- ▶ But are not we responsible for our gendered and sexist beliefs?

Aliefs/implicit biases

- ▶ In an important sense, we do not seem to be responsible for being afraid of height.
- ▶ No we should try to revise those aliefs, because it is helpful to have those kind of implicit fears.
- ▶ But are not we responsible for our gendered and sexist beliefs?
- ▶ And should not we try to revise them?

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?
- ▶ The content of aliefs seems to serve some evolutionary purpose—protect us from danger, for example.

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?
- ▶ The content of aliefs seems to serve some evolutionary purpose—protect us from danger, for example.
- ▶ No similar justification there can be for the content of *implicit biases*.

Aliefs/implicit biases

- ▶ Are Gendler's aliefs the same kind of implicit attitudes as implicit biases?
- ▶ The content of aliefs seems to serve some evolutionary purpose—protect us from danger, for example.
- ▶ No similar justification there can be for the content of *implicit biases*.
- ▶ It is interesting and puzzling that Gendler mixes together aliefs and implicit biases under the same heading. Do not you think?