39

Implicit Bias

Alex Madva

Introduction

When most people think about bias – say, racism or sexism or homophobia – they normally think about someone who openly disparages members of the group, and may act in ways that treat those members as second-class citizens. This view of bias is too narrow. It ignores the myriad ways in which we do not recognize, and may emphatically disavow, our biases toward others. This phenomenon has become known as "implicit bias." There is substantial evidence that most of us have implicit biases. Beginning in the 1980s and 1990s, researchers developed tests to detect attitudes that individuals were unable to see or unwilling to acknowledge. One of the most popular measures is the Implicit Association Test (IAT).

This test asks participants to sort words and pictures into categories as fast as possible while making as few mistakes as possible. It compares participants' ability to quickly and accurately sort items in ways that match and others that clash with common prejudices. Researchers contend that participants' speed and accuracy on these tasks indicate the strength of their associations (between, for example, black faces and negative words); they further contend that the strength of participants' associations alters how those people interact with people from different racial and social groups.

Still other tests, like the Affect Misattribution Procedure (AMP), present participants with stimuli (e.g., an image of a black face for a fraction of a second) and then assess their effects ~~of~~ on the subjects' thoughts, feelings, and actions. For more in-depth explanation of the IAT, AMP, and other implicit measures, see Brownstein (2015) and Gawronski and De Houwer (2014).

By using these tests, researchers found, for instance, that undergraduates have more difficulty associating Hispanic names (like "Juanita" and "Miguel") than non-Hispanic names ("Nicole" and "Robert") with words indicating intelligence ("brainy," "smart," etc.) (Weyant, 2005). In one especially influential 2002 study, researchers found that although many white college students explicitly endorsed anti-racist attitudes, their actions revealed racially biased ones (Dovidio, Kawakami, and Gaertner, 2002). These students claimed to be not racist, but exhibited racial biases on the IAT. In a second element of the experiment, the same students talked with one white and one black individual. During these interactions, the participants' *explicit* anti-racist attitudes best predicted whether they appeared to be friendly with black interlocutors, while their *implicit biases* best predicted nonverbal "microbehaviors." For instance, they made less eye contact, blinked more often, and sat farther away from blacks than from whites. Although their microbehaviors belied their express claims, the white participants generally thought their conversations with blacks went well, while blacks tended to think that whites were consciously prejudiced against them.

Ethics in Practice: An Anthology, Fifth Edition. Edited by Hugh LaFollette.

The findings, the researchers claim, reflect a significant fact about our society. It is "characterized by this lack of perspective … Understanding both implicit and explicit attitudes helps you understand how whites and blacks could look at the same thing and not understand how the other person saw it differently" (Carpenter, 2008, p. 36).

Widespread and frequent repetitions of these microbehaviors accumulate over time to reinforce disparities between social groups (Valian, 1998). Cortina and colleagues (2013) found that women, and especially women of color, reported more interpersonal incivility in the workplace than do men. In many cases, the incivility did not consist in overt or intentional harassment, or involve explicit reference to gender or race; rather it consisted in generic forms of rudeness, such as speaking condescendingly or interrupting a female colleague. Discourteous behavior of this sort is deeply ambiguous: since just about *everybody* interrupts and gets interrupted *sometimes*. Therefore, it is difficult to identify any particular instance of interruption as expressive of bias, as opposed to, say, misplaced enthusiasm. But because women – and especially women of color – were treated in these ways more often, they were more likely to quit their jobs.

An important aside: claims about implicit bias do not undermine claims about the way in which institutional structures impact minorities and women. That is a phenomenon more directly addressed in Chapter 57 in this volume. My point here is simply that individuals' subtle expressions of implicit bias can support and exacerbate other more explicit forms of discrimination. Moreover, it may be that implicit biases are among the obstacles that prevent us from understanding and changing unjust institutions (Madva, 2016, 2017).

For the remainder of this essay, I assume these researchers' findings reveal a significant fact about most people. With that assumption in place, I want to determine how much responsibility individuals bear for implicit bias. My answer is "some." Others disagree. They contend that holding individuals responsible or blameworthy for implicit bias is morally unfair and strategically counterproductive. I will call these theorists Exonerators. For example, Charles R. Lawrence, III writes:

> Understanding the cultural source of our racism obviates the need for fault, as traditionally conceived, without denying our collective responsibility for racism's eradication. We cannot be individually blamed for unconsciously harboring attitudes that are inescapable in a culture permeated with racism. And without the necessity for blame, our resistance to accepting the need and responsibility for remedy will be lessened. (Lawrence, 1987, pp. 325–6)

Exonerators argue that holding individuals responsible is unfair since implicit biases are unconscious and inevitable byproducts of being reared in a systemically racist world. The idea is this: individuals may be unaware of their biases, and even when made aware many individuals cannot control them. Many people contend that no one can be responsible for a condition of which they are unaware. They plausibly argue that people cannot control impulses they do not see.

Holroyd contends, however, that although we should consider these excusing conditions, people need not be consciously aware of their biases to be responsible for them (Holroyd, 2012, p. 294; see also Adams, 1985; Smith, 2005). If nothing else, accepting ignorance as an excuse would make us unduly skeptical about moral responsibility since most of us are unaware of all the factors influencing our behavior.

Although I find Holroyd's approach meritorious, I wish to challenge the Exonerators on their own terms by arguing that although awareness and control are usually sufficient for moral responsibility, many individuals are aware enough of their biases to hold them morally responsible for them. Furthermore, I think most individuals can sufficiently control their implicit biases to hold them morally responsible for them.

I think most critics err in assuming responsibility is all or nothing. I disagree. I think that responsibility is in degrees. I shall argue that this view is both plausible and useful for understanding responsibility for implicit biases.

Awareness of Implicit Bias and Discrimination: Empirical Evidence

Tony Greenwald and Mahzarin Banaji introduced the term *implicit attitudes* to counter the prevailing social-psychological assumption that most people could accurately report their beliefs and feelings about race, gender, and other social categories. They developed the IAT to measure implicit attitudes, which they defined as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (Greenwald and Banaji, 1995, p. 8). The pressing question is: *are* implicit attitudes inevitably hidden from the agent, or are they simply unidentified (or inaccurately identified)?

Initially researchers assumed that implicit attitudes were hidden from agents since subjects did not report them. They therefore inferred that everyone was wholly unaware of them. However, over a decade of research suggests this inference is too strong. Researchers have found numerous ways to reveal that disparities between overt and indirect biases are smaller than many had initially supposed. When participants separately report *both* their initial "gut reactions" and their considered "actual feelings" about social groups, these supposedly different reactions are closely correlated (Ranganath, Smith, and Nosek, 2008). In these studies, "gut feelings" are defined as the initial, spontaneous affective reactions that participants have upon first meeting or thinking about members of other social groups, where these initial reactions may be distinct from the emotions or beliefs they would endorse upon reflection (Hahn and Gawronski, 2019). And what these studies find is that most individuals are more willing to report prejudiced sentiments, and even accurately predict their biases on the IAT, if they can immediately disavow them. Participants are also more willing to report prejudiced attitudes if they are told before taking the IAT that it is an "accurate measure of racial attitudes … the closest thing to a lie detector that social psychologists can use to determine your true beliefs about race" (Nier, 2005, p. 43).

Perhaps most striking, Erin Cooley and colleagues (2015) found that simply telling participants whether their "gut feelings" did or did not reflect their "genuine" views influenced their later self-reports. Some participants were told that the negative gut feeling they may have had while looking at photos of same-sex couples reflected their "genuine attitude towards homosexuality." Those who had stronger anti-gay implicit biases were, on a subsequent questionnaire, significantly more likely to oppose gay marriage. This research thus shows how easy it is for authority figures to *legitimize* negative reactions to others, effectively transforming implicit bias into overt discrimination. This likely helps to explain the recent resurgence of explicit bigotry and misogyny in North America and Europe. However, it was heartening to see Cooley's finding that "genuine attitude" manipulation only elicited prejudiced self-reports from individuals who displayed strong implicit biases. Those without strong implicit biases were impervious to this rhetorical maneuver.

Evidence for introspective awareness of these social gut feelings is not new. In 1991, before the term "implicit attitude" was coined, Devine and colleagues found evidence for robust self-awareness of tacit prejudice among individuals who espoused anti-prejudiced ideals. This was especially obvious in research asking participants to report both how they *would react* and how they thought they *should* react. They were asked, for example, to imagine if they would be uncomfortable if "a Black person boarded a bus and sat next to you" (Devine et al., 1991, p. 819). Many thought that it shouldn't bother them, but that, in fact, it would.

Other questions revealed similar responses: would you feel "uncomfortable that a job interviewer is Black," and would you, when "seeing a Black woman with several small children," think "How typical"? A majority of people reported significant discrepancies between how they would and should respond. Most who reported such discrepancies also reported guilt or disappointment with themselves. Such findings suggest not just mere awareness of the existence and content of biased gut reactions, but a relatively rich knowledge of both their moral significance and their effects on a range of thoughts, feelings, and actions.

I take contemporary empirical research on implicit bias to reveal how we understand these purported hidden phenomena. However, I contend that Natalia Washington and Dan Kelly (2016, p. 23) misstate the evidence in claiming that: "In 1980, *no one* knew the unsettling psychological facts about implicit biases; the psychological research had not yet been done, and so today's wealth of empirical evidence simply did not exist." They thereby infer that people prior to that time were not responsible for their biases since they were not aware that they or anyone else could have such hidden biases However, it is implausible to claim no individuals *were* blameworthy in 1980 but *are* blameworthy now. It is more plausible to suggest that some individuals were *less* blameworthy then. Individuals are, or can be, introspectively aware of their biased gut reactions, and aware of discrepancies between how they think that they would and should respond in various situations. I do not mean to suggest that awareness of our biases is *easy*. To the contrary, I believe that awareness and self-knowledge are often quite difficult, and must draw on a rich store of observations and interpretations of our own and others' behavior (Madva, ~~forthcoming),~~ Indeed, perhaps none of our attitudes are directly observable just by "looking inside" our own minds (King and Carruthers, 2012), and all self-knowledge requires us to observe how we act and make inferences about ourselves based off of such observations. What this research suggests, however, is that our ability to make accurate inferences about our implicit biases is comparable to our ability to make inferences about our explicit attitudes.

If I am right, we must ask if people are morally responsible for partial and inarticulate awareness of implicit bias? One relevant bit of evidence is simply that most of us hold others responsible for their implicit biases. In Cameron, Payne, and Knobe (2010), participants assessed the responsibility of hypothetical people influenced by implicit racial bias. Some participants read about an individual who "thinks people should be treated equally, regardless of race" but "has a sub-conscious dislike for African Americans" (2010, p. 276). He tries to promote people on merit alone, but "because he is unaware of this sub-conscious dislike," he "sometimes unfairly denies African Americans promotions." Other participants read about an individual who "has a gut … dislike toward African Americans" – which he is aware of, and sincerely rejects, but has "difficulty controlling." Participants tended to judge that the individual who was aware of his dislike, but struggled to control it, was significantly more responsible and blameworthy than the one who was unaware of it altogether.

Responsibility in Degrees

Even if awareness is a necessary condition for moral responsibility, research on implicit bias (and on consciousness more generally) suggests that awareness comes in degrees, as ~~are~~ the related concepts of control, responsibility, and blameworthiness (Björnsson and Persson, 2013; Buchak, 2014; Coates and Swenson, 2013; Raz, 2010; Sinnott-Armstrong, 2013).

Although my primary focus in this essay is awareness, briefly consider control. Exonerators point out that even if individuals are aware of their implicit biases, they may not be able to control them properly. Sometimes trying to control implicit bias even exacerbates its harms (Norton et al., 2006). At issue here is what I call *local control*, the ability to directly control implicit bias, rather than *indirect control*, which involves taking steps in advance to block discrimination (e.g., anonymous reviewing), and *long-term control*, which involves debiasing one's social habits through repeated practice (Madva, 2017; Holroyd, 2012). Acknowledging that it may be difficult to control biases does not make them *uncontrollable*. I contend that although controlling implicit bias may be difficult, it is not impossible. It is impossible to control only in rare circumstances. Similarly, controlling our bladders can sometimes be difficult, but, for healthy adults, we lose control altogether only in extreme circumstances, for instance, if someone is unwell, pregnant, or overcome by intense laughter. In the same vein, control of implicit biases is also a matter of degree.

Recall that some Exonerators claim that awareness and control are necessary for moral responsibility. I disagree, although I do think awareness and control are, other things being equal, sufficient. The reasoning here is straightforward: many individuals are somewhat aware and somewhat in control of their implicit biases, and so they are somewhat responsible, and somewhat to blame. They are more responsible than they would be for purely unconscious, reflexive, or pathological behavior, but they are less responsible than they would be for explicit discriminatory behavior. When individuals' awareness of their implicit biases is not fully comprehensive or articulate – as when an implicit bias is felt but not noticed, or noticed but misinterpreted, or when it is interpreted correctly but its causal influence on judgment or action is underestimated – their responsibility is mitigated. They make the individual in question less responsible but not completely off the hook. Determining the precise extent of an individual's responsibility and blameworthiness for a specific action or omission is a complex, context-sensitive affair, analogous in some respects to the nuances and challenges of determining criminal or civil liability. My claim here is not unique to implicit biases. The same thing can be said about most moods.

Consider how we morally evaluate people's moods. Suppose Gertie is in a grumpy mood. She might not know why she is grumpy; she may not even recognize that she is grumpy. The best explanation could be stress, hunger, a headache, air pollution (Rotton et al., 1978), hitting one red light too many during the morning commute, or simply "waking up on the wrong side of the bed." She is likely also unaware of the mood's effects. Although Gertie may not notice that she is in a grumpy mood, she is in one just the same, and might, at some level, be aware of it.

However, her failure to fully notice her grumpy mood does not excuse her subsequent rudeness. We routinely hold others and ourselves responsible for the things we say and do while in a bad mood. If Gertie's mood leads her to roll her eyes, adopt a cantankerous tone, or interrupt a friend or colleague (or even a stranger), we plausibly hold her responsible. If Mordy is in a good mood because he has just received exciting news, he might fail to react with sufficient empathy and concern when he discovers Gertie has received bad news, and Gertie might reasonably hold it against him.

It's true that being in a mood can *make a difference* to responsibility and blame. Citing a bad mood as a (partial) explanation for inappropriate behavior can make the behavior appear less objectionable, somehow mitigating the severity of the offense (perhaps the behavior comes to seem less intentional or "personal"). It could be that citing a bad mood leads us to judge that the behavior is less blameworthy, or it could be that citing a bad mood leads us to shift blame from the behavior itself to the failure to restrain the behavior. In the latter case, individuals might just be responsible for letting the mood get the best of them.

Some, however, contend that mood-related misbehavior is blameless. For instance, Levy (2011, p. 245) mentions a case in which, "George's shortness with his colleagues might be excused because of the stress he has been under recently," especially if George is not fully aware of his behavior. I agree that stress can play a mitigating role, but Levy overstates its exonerating force. If we seriously entertain being one of George's colleagues, would we ordinarily take his stress to *fully* exculpate his shortness? I think not (although it might depend to some extent on how trying or traumatic the source of stress is). A graded conception of responsibility more naturally accommodates this sort of case. Learning that George has been under stress may help to make his shortness more intelligible to his colleagues, but the fact that his behavior now makes sense does not give him free license to be uncivil. On my view, George's being under stress might mitigate the severity of his offense, making him *less* responsible and blameworthy, without becoming completely off the hook for his rudeness.

The mitigating status of moods can be illuminated by considering how we offer and accept apologies for mood-influenced behavior. When George snaps at his colleague, and subsequently apologizes, he might say, "I'm sorry for being irritable. I've just been under a lot of pressure lately," or, "I just woke up on the wrong side of the bed today." How would his colleague respond in this case? Would the colleague say, "Come now, you have *nothing* to apologize for. You didn't do anything wrong." More likely, the colleague would say something like, "It's okay. Don't worry about it. I know things have been stressful for you." The apology is not out of place here, I argue, because citing a bad mood does not completely absolve one of responsibility or blame. It often has the effect of putting the person on the receiving end of the rudeness in a position to *accept* the apology, or acknowledge it some way, rather than deny the need for it altogether.

The mitigating role of moods is, of course, subject to a number of complicating factors. For one thing, it makes a difference what kind of behavior is explained by the bad mood. Being in a mood can affect an individual in many ways beyond unfriendly microbehaviors, perhaps by making them a harsher grader or a less sympathetic interviewer. What if George's stress leads not just to shortness but to verbally abusive screams, or significant property damage, or violence? If a mood leads an individual to punch someone in the face, or to deny a parole application (Danziger, Levav, and Avnaim-Pesso, 2011), then citing it might do considerably less mitigating work. Other things equal, the more serious the consequences of the behavior, the less mitigating we'll take a bad mood to be. If we are ever justified in adjusting our attributions of responsibility and blame in light of the severity of consequences, then part of the justification might lie in our (more or less explicit) knowledge that people are often better able to control themselves when the stakes are raised. For example, someone in a bad mood might be much more able, or at least more likely, to restrain his rude impulses in the presence of an armed mugger than in the presence of a close friend, or – to take an example more pertinent to implicit bias – in the presence of his bosses than in the presence of subordinates. With this understanding to hand, how should we evaluate implicit biases?

Awareness, Responsibility, and Implicit Bias

Now consider two cases that involve not (merely) bad moods but implicit biases. First, take George Yancy's rich phenomenological analysis of stepping into an elevator:

> Well-dressed, I enter an elevator where a white woman waits to reach her floor. She "sees" my Black body, though not the same one I have seen reflected back to me from the mirror on any number of occasions. Buying into the myth that one's dress says something about the person, one might think that the markers of my dress (suit and tie) should ease her tension. What is it that makes the markers of my dress inoperative? She sees a Black male body "supersaturated with meaning, as they [Black bodies] have been relentlessly subjected to [negative] characterization by newspapers, newscasters, popular film, television programming, public officials, policy pundits and other agents of representation". Her body language signifies, "Look, *the* Black!" On this score, though short of a performative locution, her body language functions as an insult. Over and above how my body is clothed, she "sees" a criminal, she sees me as a threat. Independently of any threatening action on my part, my Black body, my existence in Black, poses a threat.
>
> There is not anything as such that a Black body needs to do in order to be found blameworthy. As such, the woman on the elevator does not really see me, and she makes no effort to challenge how she sees me … she may come to judge her perception of the Black body as epistemologically false, but her racism may still have a hold on her lived body. I walk into the elevator and she feels apprehension. Her body shifts nervously and her heart beats more quickly as she clutches her purse more closely to her. She feels anxiety in the pit of her stomach. Her perception of time in the elevator may feel like an eternity … The point here is that deep-seated racist emotive responses may form part of the white bodily repertoire, which has become calcified through quotidian modes of bodily transaction in a racial and racist world … Despite how my harmless actions might be constructed within her white racialized framework of seeing the world, I remain capable of resisting the white gaze's entry into my own self-vision. I am angered. Indeed, I find her gaze disconcerting and despicable. (2008, pp. 846–47)

Drawing on W.E.B. Du Bois, Frantz Fanon, and Robert Gooding-Williams, Yancy mines this scenario to make several points. I will highlight just a few:

1 Yancy is wearing a suit and tie. It is unlikely that a white man so adorned would be perceived as a similar threat,[1] but the woman's racial attitudes bias her cognition, and prevent her from noticing or being moved by these standard markers of "respectability."

2 Yancy takes for granted that the woman's distorted perception reflects the fact that she has been bombarded with stigmatizing representations of black men in "mass media." That is, it's built into the case that her biased reaction (a range of automatic perceptual, cognitive, affective, and bodily responses) is a product of her immersion in a systemically biased world.

3 The woman might reflectively disavow her implicit biases, that is, harbor sincere egalitarian commitments and "judge her perception of the Black body" to be false.

4 But her biased reaction nevertheless constitutes, or is at least experienced as, an insult to Yancy – a tacit act of blame – which is despicable and elicits justified anger. Although Yancy is acutely aware of factors like (2) and (3), which plausibly mitigate her responsibility and blameworthiness, his moral resentment persists.

It could be that Yancy would feel less resentment and benefit psychologically if he actively concentrated on the mitigating factors, for example, by reminding himself that the woman's biases are simply byproducts of an upbringing in an unjust social reality, but the presence of these factors does not obviously make her blameless. She might not be as blameworthy as she would be if she reflectively endorsed her implicit biases, but she is more blameworthy than she would be for a mere behavioral reflex, like blinking in response to a bright light. In fact, I think that if we seriously imagine ourselves in Yancy's shoes, it is quite difficult to insist that the woman is completely unconscious of these reactions, or completely free of blame. The suggestion that her affective-bodily responses are on a moral par with behavioral reflexes, or that it is Yancy's responsibility to exercise cognitive-therapeutic techniques to reduce his stress (rather than the woman's responsibility to not act that way) strike me as morally dubious. While we need not conclude of her, or each other more generally, that we are all bad prejudiced people, it is fair to conclude that she could be, in an important sense, better than she is, and that we could be better than we are. This is an issue that pervades Kleinig's essay on the policing of minorities, Chapter 54 in this volume.

The potential for implicit biases to influence what we notice and how we respond is also evident in the following interaction described by Virginia Valian:

> A storm has damaged a large tree in the back yard, and a tree surgeon has come to look at it … As I ask the tree expert various questions about the damage and what needs to be done, I feel there is something a little odd about his responses. Finally, I realize that I am looking at him when I ask my questions, but that he is looking at J when he answers them. For his part, J is mostly looking abstractly out into space, reflecting his lack of interest in the proceedings. For the entire consultation, in fact, J is silent. I continue with my questions, and the surgeon continues to direct his answers to J. Perhaps he is riveted by J's virtuosic ventriloquism. I got the information I wanted, but I don't know what modifications I might have made – speaking louder? asking longer questions? being more assertive? – to get the tree surgeon to talk to me instead of J. I can imagine the surgeon saying to his crew afterward, "Did you see that woman? She didn't let that guy get a word in edgewise." J himself has noticed nothing, because he has been thinking about something else the whole time.~~\~~ (1998, p. 146)

In this case, my intuition is that the tree surgeon is somewhat responsible and blameworthy for failing to attend to the relevant social cues, and for acting in an oblivious, uncivil way. Unless he suffers from a visual impairment, extreme social anxiety, and so on, he knows who is asking the questions, and how to answer a person who asks a question. His social environment is giving him ample information to suggest that he should adjust his unreflective behavior, but he fails to absorb it. Valian would be warranted in resenting *him* for this behavior, rather than just, say, resenting American culture more broadly for leading the tree surgeon to develop these habits of selective attention (although she could reasonably resent American culture, too). In this case, the tree surgeon's bias may not lead him to express any sort of negative affect, as did the woman on the elevator, but it does lead him to discount or ignore what's right in front of him, and culpably so. Part of the explanation for this, I believe, is the implausibility of supposing that the tree surgeon is *completely unaware* of what he's doing. Suppose Valian had said something explicitly about his failure to make eye contact. He might have reacted with defensive hostility or denied that he meant any ill will, but would he really have had no clue what she was even *referring to*?

It is not just that we can trace things back to some prior moment in which the individual should have reflected upon things and decided to form better social habits; there is a kind of awareness operative *at the time*, as the conversation is unfolding, which puts the individual in contact with the relevant feature of the situation, and on the hook for acting appropriately. To spin this point in more forward-looking terms, one reason to make a lot of hay out of the sort of first-personal awareness that individuals seem to have of their implicit biases is that this awareness *presents an opportunity for intervention*. Implicit biases aren't just coloring our thoughts, perceptions, and actions from behind the locked door of the unconscious, but are themselves palpably present (or at least accessible) to awareness. This first-personal, in-the-moment awareness of our biased thoughts, feelings, and actions opens up a distinctive set of opportunities for us to do better.

Conclusion

Of course, it may or may not be productive for Valian or Yancy to say something accusatory in such situations. In this vein, Exonerators argue that holding individuals responsible for implicit bias would be not just unfair but also strategically ill-advised. While I agree that we should not necessarily saddle individuals with "-ist" labels that portray them as horrible people for possessing and expressing implicit biases, it is a mistake to conflate sanctimonious name-calling with the view that implicit bias is often worthy of blame, broadly construed. Blame is not so blunt an instrument. We can acknowledge the failings of others and ourselves to live up to our commitments without calling the sincerity of those commitments into question. In many cases, we can insist that individuals bear a legitimate degree of responsibility and blame, even if they lack perfect awareness of what they do. If it is ever strategically unwise to lay blame, then the upshot is not to jettison implicit bias from the sphere of moral responsibility; the upshot is to take great care in locating it properly within that sphere. Needless to say, I am not suggesting that the most effective way to combat systemic discrimination and oppression is simply to stamp out individuals' biased microbehaviors. We should combat systemic ills by changing the system. In addition to thinking about how the system needs to change, we cannot forget who needs to change it.

Note

[1] It must be acknowledged that many women live with experiences of vulnerability that partly inform what's going on in this encounter. It's possible, for example, that this individual's past experiences with harassment would make her uncomfortable alone with any man in this situation. However, if we stipulate, e.g., that they are riding the elevator in an otherwise-crowded, publicly accessible building in the middle of the day in an extremely low-crime area, etc., it is reasonable to expect that many white women would not feel equally uncomfortable around a similarly dressed white man.

References

Adams, Robert Merrihew. 1985. "Involuntary Sins." *The Philosophical Review* 94 (1): 3–31. https://doi.org/10.2307/2184713.

Björnsson, G., and K. Persson. 2013. "A Unified Empirical Account of Responsibility Judgments." *Philosophy and Phenomenological Research* 87 (3): 611–39.

Brownstein, Michael. 2015. "Implicit Bias." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2017. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2017/entries/implicit-bias/.

Buchak, L. 2014. "Belief, Credence, and Norms." *Philosophical* 169 (2): 285–311.

Cameron, C.D., B.K. Payne, and J. Knobe. 2010. "Do Theories of Implicit Race Bias Change Moral Judgments?" *Social Justice Research* 23 (4): 272–89.

Carpenter, S. (April/May. 2008. "Buried Prejudice." *Scientific American Mind* 35.

Coates, D.J., and P. Swenson. 2013. "Reasons-Responsiveness and Degrees of Responsibility." *Philosophical Studies* 165 (2): 629–645.

Cooley, Erin, B. Keith Payne, Chris Loersch, and Ryan Lei. 2015. "Who Owns Implicit Attitudes? Testing a Metacognitive Perspective." *Personality and Social Psychology Bulletin* 41 (1): 103–15. https://doi.org/10.1177/0146167214559712.

Cortina, Lilia M., Dana Kabat-Farr, Emily A. Leskinen, Marisela Huerta, and Vicki J. Magley. 2013. "Selective Incivility as Modern Discrimination in Organizations: Evidence and Impact." *Journal of Management* 39 (6): 1579–605. https://doi.org/10.1177/0149206311418835.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. 2011. "Extraneous Factors in Judicial Decisions." *Proceedings of the National Academy of Sciences* 108 (17): 6889–92. https://doi.org/10.1073/pnas.1018033108.

Devine, P.G., M.J. Monteith, J.R. Zuwerink, and A.J. Elliot. 1991. "Prejudice with and without Compunction." *Journal of Personality and Social* 60 (6): 817–30.

Dovidio, John F., Kerry Kawakami, and Samuel L. Gaertner. 2002. "Implicit and Explicit Prejudice and Interracial Interaction." *Journal of Personality and Social Psychology* 82 (1): 62–8. https://doi.org/10.1037//0022-3514.82.1.62.

Gawronski, Bertram, and Jan De Houwer. 2014. "Implicit Measures in Social and Personality Psychology." In *Handbook of Research Methods in Social and Personality Psychology*, edited by Charles M. Judd and Harry T. Reis, 2nd edn, 283–310. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511996481.016.

Greenwald, A.G., and M.R. Banaji. 1995. "Implicit social cognition: Attitudes, self-esteem, and stereotypes." *Psychological Review* 102: 4–27.

Hahn, Adam, and Bertram Gawronski. 2019. "Facing One's Implicit Biases: From Awareness to Acknowledgment." *Journal of Personality and Social Psychology*, 116 (5): 769–94.

Holroyd, Jules. 2012. “Responsibility for Implicit Bias.” *Journal of Social Philosophy* 43 (3): 274–306. https://doi.org/10.1111/j.1467-9833.2012.01565.x.

King, M., and P. Carruthers. 2012. “Moral responsibility and consciousness.” *Journal of Moral Philosophy* 9 (2): 200–28.

Lawrence, Charles R., III. 1987. “The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism.” *Stanford Law Review* 39 (2): 317–88. https://doi.org/10.2307/1228797.

Levy, N. 2011. “Expressing Who We Are: Moral Responsibility and Awareness of Our Reasons for Action.” *Analytic Philosophy* 52 (4): 243–61.

~~Madva, Alex. Forthcoming. “Social Psychology, Phenomenology, and the Indeterminate Content of Unreflective Racial Bias.” In *Race as Phenomena*, edited by Emily S. Lee. Rowman and Littlefield.~~

Madva, Alex. 2016. “A Plea for Anti-Anti-Individualism: How Oversimple Psychology Misleads Social Policy.” *Ergo, an Open Access Journal of Philosophy* 3 (27): 701–28. https://doi.org/10.3998/ergo.12405314.0003.027.

Madva, Alex. 2017. “Biased against Debiasing: On the Role of (Institutionally Sponsored) Self-Transformation in the Struggle against Prejudice.” *Ergo, an Open Access Journal of Philosophy* 4 (6): 145–79. http://dx.doi.org/10.3998/ergo.12405314.0004.006.

Nier, J.A. 2005. “How Dissociated Are Implicit and Explicit Racial Attitudes? A Bogus Pipeline Approach.” In *Group Processes and Intergroup Relations*, 8: 39–52.

Norton, M.I., S.R. Sommers, E.P. Apfelbaum, N. Pura, and D. Ariely. 2006. “Color Blindness and Interracial Interaction Playing the Political Correctness Game.” *Psychological Science* 17 (11): 949–53.

Ranganath, Kate A., Colin Tucker Smith, and Brian A. Nosek. 2008. “Distinguishing Automatic and Controlled Components of Attitudes from Direct and Indirect Measurement Methods.” *Journal of Experimental Social Psychology* 44 (2): 386–96.

Raz, J. 2010. “Being in the World.” *Ratio* 23 (4): 433–52.

Rotton, J., T. Barry, J. Frey, and E. Soler. 1978. “Air Pollution and Interpersonal Attraction.” *Journal of Applied Social Psychology* 8 (1): 57–71.

Sinnott-Armstrong, W. 2013. “Are Addicts Responsible?. Addiction and Self-Control: Perspectives From.” *Philosophy, Psychology, and Neuroscience*, 122–42.

Smith, Angela M. 2005. “Responsibility for Attitudes: Activity and Passivity in Mental Life.” *Ethics* 115 (2): 236–71. https://doi.org/10.1086/426957.

Valian, Virginia. 1998. *Why So Slow? The Advancement of Women*. Cambridge, MA: MIT Press.

Washington, N.T., and D. Kelly. 2016. “Who’s Responsible for This? Moral Responsibility, Externalism, and Knowledge about Implicit Bias.” In *Implicit Bias & Philosophy: Volume 2: Responsibility, Structural Injustice, and Ethics*, edited by M. Brownstein and J. Saul. Oxford: Oxford University Press.

Weyant, James. 2005. “Implicit Stereotyping of Hispanics: Development and Validity of a Hispanic Version of the Implicit Association Test.” *Hispanic Journal of Behavioral Sciences* 27 (3): 355–63. https://doi.org/10.1177/0739986305276747.

Yancy, G. 2008. “Elevators, social spaces and racism: A philosophical analysis.” *Philosophy & Social Criticism* 34 (8): 843–876.