Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Ariella Shulman April 8, 2025

Agency and Automaticity: Examining implicit prejudice through the lens of Deweyan habits

by

Ariella Shulman

Jessica Wahman

Adviser

Philosophy

Jessica Wahman

Adviser

John Stuhr

Committee Member

Dr. Jessica Barber

Committee Member

2025

Agency and Automaticity: Examining implicit prejudice through the lens of Deweyan habits

By

Ariella Shulman

Jessica Wahman

Adviser

An abstract of

a thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Arts with Honors

Philosophy

2025

Abstract

Agency and Automaticity: Examining implicit prejudice through the lens of Deweyan habit
By Ariella Shulman

In this thesis, I claim that people have the capacity to exercise control over their own implicit prejudice. I argue that implicit prejudice, which is described in psychological literature as an automatic process, should not be taken as a deterministic threat to agency because the fact of its automaticity does not necessarily preclude deliberate control on the part of the agent. Agency and automaticity are not incompatible with each other, despite automaticity posing a legitimate challenge to immediate and direct forms of agency. However, the current landscape of interventions that are designed to help people combat their own implicit prejudice are ineffective, in part due to their treatment of agency as a subjective and individualistic pursuit. Using John Dewey's reconstruction of agency as degrees of flexibility, I show that agency is a social and interdependent process through which individuals may gain, in degrees, the ability to intelligently deliberate between possibilities. Through this lens, I evaluate how implicit prejudice is a socially constructed and transmitted phenomenon, and how any program aiming to reduce implicit prejudice in individuals should locate its intervention in the relationship between an individual and their social environment.

Agency and Automaticity: Examining implicit prejudice through the lens of Deweyan habits

By

Ariella Shulman

Jessica Wahman

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences

of Emory University in partial fulfillment

of the requirements of the degree of

Bachelor of Arts with Honors

Department of Philosophy

2025

Acknowledgments

Firstly, I owe a great deal to Professor Wahman, who has been a trusted source of wisdom, knowledge, and guidance from my very first semester at Emory until my very last. I am also immensely grateful to Professor Stuhr for his knowledge of pragmatism and willingness to share this knowledge, and to Professor Barber, whose exceptional teaching sparked conversations that inspired me to write this thesis. In addition, I am thankful to Carla Freeman, Shiv Sharma, and everyone at the Fox Center for providing such a supportive and motivating community of fellow writers. Finally, thank you to Victoria Keane for always inspiring me to search for answers, whether in the pages of a book or through the lens of a telescope.

Table of Contents

Introduction……..1

I. Implicit Prejudice…….5

History and Terminology…….7

Current Research……....10

II. Agency……..15

The Free Will Debate…….15

Dewey and the Self……20

Flexible Habits……...26

III. Automaticity……32

Old Ideas……33

New Environments…….38

IV. Improving Flexibility……...41

A Pathology of Goodness……..42

Adaptive Agents…….44

Conclusion…….46

Bibliography……..51

Introduction

"After my election, there was talk of a post-racial America. Such a vision, however well-intended, was never realistic. Race remains a potent and often divisive force in our society … But laws alone won't be enough. Hearts must change. It won't change overnight. Social attitudes oftentimes take generations to change."

— Barack Obama, 2016[1]

"A social revolution may effect abrupt and deep alterations in external customs, in legal and political institutions. But the habits that are behind these institutions and that have, willy-nilly, been shaped by objective conditions, the habits of thought and feeling, are not so easily modified. They persist and insensibly assimilate to themselves the outer innovations … The force of lag in human life is enormous."

— John Dewey, 1922[2]

As President Barack Obama stood to address the nation one final time before handing the reins of power to Donald Trump, he offered a solemn message to anyone who thought the presence of a Black man in the Oval Office meant the end of racism in America. Obama recognized how shifting legal institutions are insufficient to alter the undercurrent of racial prejudice coursing through American culture. "Hearts must change," he told the country, and his words are a poignant reminder of how far our society is from the egalitarian utopia we might imagine it to be.

---

[1] "President Obama's Farewell Address." n.d. *The White House*. Accessed March 12, 2025. https://obamawhitehouse.archives.gov/node/360231.

[2] Dewey, John. 1922. *Human Nature and Conduct: An Introduction to Social Psychology*. Henry Holt And Company, 108.

Nearly a century earlier, in 1922, American philosopher John Dewey wrote about human behavior in a remarkably similar way. Dewey was a pragmatist who believed that genuine moral inquiry could not be achieved unless philosophy learned to use human nature as a starting point for ethics. His theory of habits, which encompasses the way that humans learn, grow, and interact with the world as social beings, was instrumental to his theories of social and political progress. Dewey recognized how habits can resist a changing external environment while continuing to pull human activity in a different direction. Like Newton's forces of gravity and angular momentum that keep an object in orbit, Dewey's account of habit also describes a propulsive force: "The force of lag in human life is enormous." In 1922, Dewey was speaking to an America that looked vastly different from, and yet remarkably similar to, the America of today. Since then, social revolutions like the Civil Rights Movement catalyzed monumental shifts in American legal and political institutions. America became more equal on paper. But just as Dewey observed, the underlying habits of racial and gendered inequality were not so easily forgotten, and as America has transitioned into the 21st century, the legacy of white male supremacy continues to operate.

One of the more insidious forms of hatred that permeates American culture today is implicit prejudice—a form of prejudice that resists self-identification, operates below the level of awareness, and contrasts with explicitly-held beliefs, all while continuously shaping behavior. As an automatic process, implicit prejudice can be thought of as the brain on "autopilot:" highly efficient, but not easily modifiable. Neuropsychological research has revealed the stubbornness of implicit prejudice, how it seeps into the crevices of our cognition and gets stuck there. It may be shocking to have revealed levels of implicit prejudice towards a group of people that one genuinely never imagined themselves having. Maybe a person thinks of themselves as valuing

equality for all people, until confronted by the realization that prejudice lurks beneath the surface of their awareness. How does this change people's self-perception, and more importantly, what can they do with this information if they want to change?

If America wants to truly become more egalitarian, we need to address the problem of implicit prejudice. But can a person have power over something so hidden from their own consciousness? The answer depends on how one understands that word: *power*. The way that many theorists and psychologists define power could be detrimental for the problem of reducing implicit prejudice, framing it as inevitable in modern society or arguing that any attempts at intentional change are futile. With my thesis, I aim to prove these theories wrong, and to show that we do have power over implicit prejudice, more than we may think. This power comes in the form of long-range indirect agency. I argue that implicit prejudices can be brought within the scope of our control, but only if we understand control, and implicit prejudice itself, through a different lens. Using John Dewey's theory of humans as systems of habits, I argue for a socially-embedded understanding of implicit prejudice that accounts for change in the form of cultivating flexible habits.

As I pursue this question, I am specifically interest in one subset of the population: those who believe themselves to be unprejudiced (in thought and action), yet who find themselves having some level of prejudice that exists in contradiction to their explicitly held values and beliefs. At the heart of this situation lies a contradiction between implicit motivations and explicit values, which the individual desires to resolve by reducing their own levels of implicit prejudice. It is this population that I am interested in exploring in this thesis, in part because my own experiences closely mirror this kind of situation. This narrow framing excludes, for example, a) people who secretly harbor explicit prejudice, yet who hide these beliefs for the

purposes of social approval, and b) people who do *not* have a strongly held desire to reduce their own levels of prejudice, be they implicit or explicit. As such, I will not discuss the process of getting individuals to *care about* the problem of implicit prejudice, as this level of psychological commitment is an already-present assumption in my analysis. Rather, I am interested in the question of what people can *do* about their own implicit prejudice, once they become aware of it and have a desire to change it. I am also limiting my discussion to Western, and specifically American populations, because I am interested in the way that American culture contributes to the formation of prejudice as well as its implications for the way people think about and exercise agency.

In Part I, "Implicit Prejudice," I introduce the phenomenon of implicit prejudice and offer a brief account of its development as a psychological theory. I also examine four of the most common approaches to helping people change their own implicit prejudice and identify the benefits and drawbacks of each approach. Looking at the ineffectiveness of most interventions leads me to my main question: is it possible for people to change their own implicit prejudices, and if so, how? In Part II, "Reconstructing Agency," I begin by reviewing the contemporary free will debate to understand how implicit prejudice fits—or rather, does not fit—into traditional theories of control. I turn instead to John Dewey's theory of humans as systems of habits, showing that it is the most useful framework for addressing both aspects of the threat that implicit prejudice poses to control. In Part III, "Social Origins of Prejudice," I use Dewey's theories of custom and shared habits to understand how implicit prejudice forms in the collective minds of a society. Looking at language, media, and schooling, I will analyze how prejudice is transmitted and communicated across groups. In Part IV, "Improving Flexibility," I offer an alternative framework by which to design programs that aim to help people reduce their own

implicit prejudice. Informed by Dewey's notion of agency as degrees of flexibility, I will argue that fixating on becoming unprejudiced, as a rigid end, is hindering progress in the fight against implicit prejudice. Instead, we should prioritize flexibility among individuals in order to empower them to overcome prejudice through deliberation and habitual practice.

I. Implicit Prejudice

In my junior year of university, I came face to face with my own implicit prejudice for the first time. I was a student in Psychology of Prejudice, a class aimed at uncovering the psychological and neurological processes underlying prejudice and discrimination. As a philosophy major with an interest in psychology, I was already deeply familiar with notions of prejudice, institutional discrimination, and cultural hierarchies. I had an idea that as a white woman in higher education I was, to some extent, complicit in racial and socioeconomic power structures. Likewise, I knew from lived experience how gender discrimination and implicit misogyny can permeate a space and affect sociopolitical norms. But up to that point, I could only conceptualize these ideas as intangible, vague mental constructs. And if someone had asked, I would have described myself as thoroughly unprejudiced. That was, until I took the Implicit Association Test (IAT)[3] for Race as part of an assignment for *Psychology of Prejudice*.

The IAT is one of several tests aiming to quantify levels of implicit prejudice. Through a series of selection prompts, the IAT scores participants based on how quickly they form associations between social objects (e.g. Black men) and evaluations (e.g. good, bad, trustworthy, incompetent). The test's rapid-fire nature is intended to minimize introspection and strategic response, with the goal of measuring genuine automatic evaluations. As I clicked

[3] "About the IAT." n.d. Accessed March 12, 2025. https://implicit.harvard.edu/implicit/iatdetails.html.

through the prompts, I was too concentrated on carefully following the instructions to derive any sense of what the outcome might be. Finally, I received my results: "Slight automatic preference for European American compared to African American." I was also shown a table of score distributions: 17% of participants got the same result as I did. 51% demonstrated "moderate" or "strong" automatic preference for European American, while 18% showed no preference either way and only 14% showed any level of automatic preference for African American.

These results, while not shocking, did leave me with a feeling of unease. I was made uncomfortable by the fact that my levels of implicit racial prejudice did not match the way I would describe myself. Either I had been lying to myself and the people around me, I thought, pretending to be more egalitarian than I actually was, or there was some hidden part of myself that retained preference toward people of my own skin color—hidden, even, to myself. Research has shown that negative affect, including defensiveness, avoidance, anxiety, and frustration, is common among participants who take the IAT but consider themselves proponents of equality.[4] Regardless of how unprejudiced one may take themselves to be, the IAT will often reveal a conflicting narrative that is difficult for participants to reconcile. What I asked myself then, and what I will interrogate in the following chapters, is whether this feeling of helplessness in the face of learning about one's implicit prejudice is indicative of an absolute lack of control over implicit prejudices. I endeavor to show that this is not the case; we do have opportunities to understand where these implicit prejudices come from and how we can change them. I will argue that, by reconstructing assumptions about agency and behavioral modification, there is a

[4] Howell, Jennifer L., Sarah E. Gaither, and Kate A. Ratliff. 2015. "Caught in the Middle: Defensive Responses to IAT Feedback Among Whites, Blacks, and Biracial Black/Whites." *Social Psychological and Personality Science* 6 (4): 373–81. https://doi.org/10.1177/1948550614561127.

defensible responsibility for people to change their own implicit prejudices and strive towards more egalitarian habits.

History and Terminology

The IAT was created in 1998 by researcher Anthony Greenwald and his colleagues who sought to create a test that could quantify implicit prejudice without relying on participants' self-reporting.[5] Just three years earlier, Greenwald had published a paper with Mahzarin Banaji called "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes."[6] This paper represents the first instance of scientific literature in which implicit prejudice (then called "implicit attitudes") was named and given an operational definition: "Implicit attitudes are introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects."[7] For Banaji and Greenwald, the key aspect of implicit attitude/prejudice is that it evades self-identification while continuing to affect behavior towards others.

A number of synonyms exist, in popular culture and scientific literature, to describe the phenomenon of automatic evaluations of social objects. I will use implicit prejudice because I understand it to be the most accurate and useful term for this phenomenon. The etymology of "prejudice," derived from the Latin *praejudicium* ('in advance' + 'judgment'), points to its function: to judge (i.e. assign a moral quality) prior to intelligent inquiry (i.e. deliberate analysis of objective facts and relations). Implicit prejudice, once shaped by various cultural factors,

---

[5] Self-reporting of implicit attitudes is notoriously invalid due to the Social Desirability Response Bias: a tendency to respond consistently with social norms in order to appear unprejudiced. See: Nosek, Greenwald, and Banaji, (2005).

[6] Greenwald, A. G., and M. R. Banaji. 1995. "Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes." *Psychological Review* 102 (1): 4–27. https://doi.org/10.1037/0033-295x.102.1.4.

[7] Ibid., 8.

becomes automatic at the time of activation. It alters one's perception by creating a mental schema of social categories, and their associated traits, before a specific interpersonal interaction occurs. The precise nature of these schema is primarily colored by whether or not someone is a member of our ingroup. Ingroup formation is constructed through our membership in social institutions and networks: family, school, workplace, congregation, social media. The term "prejudice" is generally applied normatively to describe social situations, so it captures both the automaticity and intrinsic sociality of the phenomenon. Related terms, such as "bias" and "attitudes," are not as useful due to the broad scope of their meanings. Though "implicit bias" is common in popular discourse—and favored by the American Psychological Association[8]—bias can extend towards or against non-human animals, inorganic objects, and mere concepts and ideas; "prejudice" generally refers to social groups and their members. The word "discrimination" is also frequently used but should be carefully distinguished as a behavioral product of prejudice, both implicit and explicit.[9]

In the following decades, as social psychology grew as a field, more definitions for implicit prejudice were created. For a poignant example, see this definition by Mary Kite and Bernard Whitley published in 2016:

> Implicit prejudice refers to intergroup stereotypes and attitudes that are activated in memory when the person encounters a member of an outgroup without the person being

[8] https://www.apa.org/topics/implicit-bias

[9] Discrimination, at least on the individual level, is the action that occurs when a person acts on their prejudice, such as denying service to a person because of their skin color. It refers to the behavior that results from prejudice. Once a thought crosses into the realm of behavior and becomes actualized, controlled processing takes over, so it is unlikely that psychologists would identify discrimination as implicit. Discrimination also happens on the structural level. Institutional practices such as redlining, unequal lending, voter suppression, and disenfranchisement are some examples of structural discrimination. These constitute a form of oppression against an entire social group via institutional norms. Such norms are upheld and enforced by individuals but extend beyond the scope of a single actor, blurring the line between individual and structural discrimination.

> aware that the activation has occurred … Implicit prejudices lie dormant until an event occurs—such as encountering a member of an outgroup—that activates the prejudice.[10]

Kite and Whitley added elements of memory and awareness to expand on the notion of inaccurate self-identification and also proposed the dormant-active model of implicit prejudice. In general, there is no singular consensus between researchers for how to define implicit prejudice. Yet all definitions seem to have one crucial thing in common, either stated in the definition or subtly implied: that implicit prejudice is different from explicit prejudice due to the fact that implicit prejudice is an automatic process.

What is an automatic process? In the 1970s, amid a growing curiosity about how brains work coupled with improved neuroimaging technology, new terms emerged to distinguish different types of mental tasks. The dual-process theory, popularized by Daniel Kahneman in his book *Thinking, Fast and Slow*,[11] posits that there are essentially two types of brain processes: one is fast, automatic, and unconscious; another is slow, deliberative, and conscious. The former requires little directed attention and can be thought of as a person going on "autopilot:" speaking one's native language or recognizing friendly faces.[12] The latter necessitates conscious attention: solving a tricky math problem or making a difficult decision at work. Another key distinction is that automatic processing, unlike controlled processing, is very difficult to intentionally modify. Although automation is evolutionarily advantageous since it allows humans to streamline behavior and efficiently prioritize cognitive resources, it can also be prone to habitual errors and

---

[10] Kite, Mary E., Bernard E. Whitley, and Lisa S. Wagner. 2022. "Psychology of Prejudice and Discrimination." *Routledge & CRC Press*. 8. https://www.routledge.com/Psychology-of-Prejudice-and-Discrimination/Kite-WhitleyJr-Wagner/p/book/9780367408176.

[11] Daniel Kahneman. 2011. *Thinking, Fast and Slow*. Macmillan.

[12] Automatic processes are distinct from reflexes (e.g. blinking), which are rapid behavioral responses occurring in the neural pathways of the spinal cord. Reflexes rarely, if ever, reach the level of cortical processing. As a result, opportunities for conscious override of reflexes (i.e. keeping your eyes open when you sneeze) are minimal.

biases.[13] The integration of research on prejudice with that of automatic processing in the 1990s was integral to the way that social psychologists currently understand implicit prejudice as an automatic process.

Current Research

At this point, several decades of research have been dedicated towards the stated goal of helping individuals to reduce their own implicit prejudice. This research falls into four main categories: awareness, change-based, control-based, and pharmaceutical interventions.

Awareness interventions focus on increasing individuals' recognition of their own implicit prejudices. The primary goal is to help people identify unidentified judgements and associations that may influence their behavior, thereby enabling them to make more informed decisions in social interactions. These interventions often employ IATs as diagnostic tools, with participants encouraged to reflect on how the test results point to implicit judgments that affect behavior. The intended outcome is that awareness will foster a commitment to change, motivating individuals to keep engaging in efforts to reduce their own implicit prejudice. However, research has shown mixed results. While increased awareness can lead to a heightened understanding of one's own implicit prejudice, this realization does not always translate into behavioral change. In some cases, awareness may even provoke defensiveness or rationalization, where individuals justify their prejudices instead of working to modify them. In a notable study, participants who learned of their implicit prejudices sometimes reported a sense of inevitability

[13] Payne, B. K., & Gawronski, B. 2010. A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*, 1-15. The Guilford Press.

regarding their responses, thus undermining the potential for transformation.[14] Moreover, the impact of these interventions tends to be short-lived. Without a supportive framework for sustained reflection and action, the momentary insights generated by IAT feedback can dissipate. Despite these challenges, awareness-based approaches have become widespread in academic and organizational settings, where they serve as a starting point for broader diversity and inclusion initiatives.[15] Ultimately, while awareness-based interventions have succeeded in sparking conversations about implicit prejudice, their effectiveness in driving long-term behavioral change remains limited.

Change-based interventions are designed to directly modify the cognitive and affective structures underlying implicit prejudice. These interventions aim to reshape automatic associations and recalibrate the evaluative processes that lead to prejudiced responses. A common strategy in this category is counter-stereotypic training,[16] where individuals are repeatedly exposed to examples that contradict common stereotypes. For instance, a training module might involve presenting images and narratives of counter-stereotypical role models, such as women in leadership positions or racial minorities excelling in academic and professional arenas. The intended goal is to create new associative links that weaken the power of pre-existing associations. Some studies have demonstrated that repeated exposure to counter-stereotypic information can result in temporary reductions in IAT scores. The premise is that through sustained practice, individuals learn to automatically associate positive attributes with

---

[14] Devine, P. et al. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6), 1267-1278.

[15] For example, see advertisements for "unconscious bias training" and similar programming on Emory University's sites for the Office of Equity and Inclusion (https://diversity.emory.edu/) and Inclusive Pedagogy (https://cfde.emory.edu/programs/teaching/inclusive-pedagogy/index.html)

[16] "Implicit Bias | Change-Based Interventions." *Stanford Encyclopedia of Philosophy.* Accessed March 12, 2025. https://plato.stanford.edu/entries/implicit-bias/#ChanBaseInte.

previously stereotyped groups, thereby diminishing the strength of their implicit biases. However, the durability of these effects is often questioned. While short-term improvements in implicit associations are observed, many interventions struggle to produce long-lasting change. Participants may revert to ingrained biases over time if the new associations are not continually reinforced through ongoing practice or changes in the broader social environment.

Control-based interventions focus less on the content of implicit prejudice and more on how an individual can respond during moments of activation. These approaches rest on the premise that individuals may possess implicit prejudices but can exert conscious control to prevent these prejudices from influencing their behavior. Techniques in this category often draw on cognitive-behavioral strategies and mindfulness training to bolster self-regulatory capacities.[17] For example, a control-based intervention might involve training participants to recognize cues that trigger prejudiced responses and then employ cognitive reframing techniques to neutralize these impulses. In some cases, interventions include role-playing exercises and real-time feedback mechanisms that help individuals practice inhibition of stereotypic responses—although, notably, these interventions always take place in a carefully controlled laboratory setting. Evidence suggests that improving self-regulatory skills, such as inhibition and focused attention, may reduce the likelihood of acting on implicit prejudices. Studies have found that individuals who engage in mindfulness practices demonstrate a greater capacity to delay automatic responses and make more considered decisions in intergroup contexts. However, the effectiveness of these interventions can vary widely across different individuals and settings. A significant challenge is that self-regulatory resources are finite; under conditions of cognitive load or stress, individuals may be less able to inhibit biased responses. Moreover, while these

[17] "Implicit Bias | Control-Based Interventions." *Stanford Encyclopedia of Philosophy.* Accessed March 12, 2025. https://plato.stanford.edu/entries/implicit-bias/#ContBaseInte.

interventions can help mitigate the expression of prejudice in specific situations, they may not fundamentally alter the underlying implicit associations. Thus, while control-based approaches can lead to improvements in behavioral outcomes, their long-term impact on the deep-seated cognitive structures that produce bias remains limited.

Pharmaceutical interventions are a more recent and controversial approach to addressing implicit prejudice. These interventions involve the use of psychopharmacological agents to modulate neural mechanisms thought to underlie implicit prejudice. The intended goal is to reduce the neural reactivity that may contribute to automatic negative evaluations of certain groups. The most prominent example is a drug called propranolol, a beta-blocker commonly used to treat hypertension and anxiety. By dampening amygdala activity and blocking beta-adrenergic receptors, which are involved in the body's fight-or-flight response, propranolol appears to reduce the intensity of negative emotional response correlated with encountering outgroup members that may be perceived as threatening.[18] The rationale behind pharmaceutical interventions is grounded in the assumption that implicit prejudice is largely a neurochemical process. While early research has shown temporary improvement in IAT scores, experiments have largely been conducted in controlled laboratory settings with small sample sizes, and the success of such experiments remains contested. The transient nature of the effects observed in experimental contexts raises questions about the practicality of such interventions for long-term prejudice reduction. Additionally, ethical concerns regarding the use of drugs to alter attitudes and behaviors, particularly in relation to issues of autonomy, consent, and side effects remain prevalent.

[18] Terbeck, S., Kahane, G., McTavish, S. et al. 2012. Propranolol reduces implicit negative racial bias. *Psychopharmacology* (222), 419–424. https://doi.org/10.1007/s00213-012-2657-5

While all four approaches have their merits, it is evident that no one approach is sufficient for people to create lasting and meaningful reductions of their own implicit prejudice. Awareness interventions can prompt self-reflection but often fail to translate into behavioral change. Change-based approaches attempt to reconfigure automatic associations but require sustained reinforcement to be effective. Control-based strategies offer tools for managing prejudice in the moment but rely on cognitive resources that may be inconsistent under stress. Pharmaceutical interventions provide insight into the neurobiological aspects of prejudice but raise ethical concerns and lack long-term viability. The ineffectiveness of most interventions is further exacerbated by the "rebound effect,"[19] a psychological phenomenon where efforts to draw awareness to implicit prejudice can unintentionally amplify levels of prejudice in group settings. As a result, the majority of attempts at intentionally changing implicit prejudice in individuals either dissipate too quickly to be useful or inadvertently create more harm. Where does this leave us, as a society seeking to bolster social equality through the reduction of implicit prejudice? Does the solution lie in better laboratory equipment, harsher legal codes, or stricter institutional norms? Or is it up to individuals to find their own way forward, becoming unprejudiced through sheer willpower? The answer, I will show, lies in how we understand agency and a person's ability to control their own behavior. With a clear definition of agency, the work of understanding implicit prejudice and people's power in changing it can be more intelligently pursued.

[19] Shen, Z., Hunt, J. 2008. The stereotype rebound effect: Universal or culturally bounded process?, *Journal of Experimental Social Psychology* (44): 3, 489-500. https://doi.org/10.1016/j.jesp.2007.07.010.

II. Agency

To uncover if and how people can meaningfully alter their own implicit prejudice, I will turn to the concept of agency and intentional control. Whether, and in what capacity, human beings can exercise agency in general is the subject-matter of the contemporary free will debate, which consists of three main positions: hard determinism, libertarianism, and compatibilism. Hard determinists affirm determinism and reject free will; libertarians maintain that free will exists, either by transcending determinism or through the fact of indeterminism; and compatibilists say that freedom is found in acting according to one's own desires, even if those desires are determined, so free will and determinism are not mutually exclusive. In this chapter, I will analyze these three positions to understand how implicit prejudice might be understood according to their respective frameworks. As I will show, each position obfuscates a clear response to the problem of implicit prejudice, and none produce a scenario in which people can be held responsible for their own implicit prejudice or attempt to meaningfully change it. What is needed is a different conception of agency that diverges from the traditional free will debate. I will show that John Dewey's notion of the self as a system of habits provides a better way to think about agency, one that locates decision-making in the interdependent relationship between the self and the environment and allows for authentic, yet contingent, agency produced through flexible habits.

The Free Will Debate

A common way to conceptualize free will is as the ability to make choices that are "up to us." If someone's actions are up to them, then that person can be said to have free will. If their actions are up to someone or something else, then they do not have free will. The leading

positions in the contemporary free will debate can be demarcated by how they define an action as being "up to us." Both hard determinists and libertarians believe that choices are up to us if, and only if, they are uncaused. To be uncaused, a choice must originate solely from the agent and not from antecedent conditions, so determinism is incompatible with free will. Hard determinists, as the name suggests, affirm determinism by arguing that all choices are caused by antecedent conditions, thus concluding that no choices are up to us. As a result, hard determinists deny any true agency whatsoever to individuals. On the other hand, libertarians reject determinism by framing agents themselves as the uncaused causes of their own actions, making these actions up to the agent. Libertarians afford agency to individuals by arguing that individuals always (barring physical obstacles) have the opportunity to choose between alternative possibilities. Finally, unlike hard determinists and libertarians, compatibilists diverge from the premise that free will is incompatible with determinism. Generally, compatibilists argue that an action is up to us if we act according to the way we desire to act, regardless of if our desires are caused by antecedent conditions. The acting itself, compatibilists argue, is still in our hands so long as we have acted in alignment with our ultimate desires for different paths of action.

For hard determinists, the prospect of being held responsible for reducing one's own implicit prejudice would be a nonstarter. Hard determinism says that change, when it happens, is necessarily determined by antecedent conditions. Choices are made according to the unfolding of external causes, such as laws of nature or God's will, with no room left over for human agency. In the absence of genuine responsibility, consciously reducing one's own implicit prejudice is not a real possibility. There are several popular neuroscientists who follow the theory of hard determinism to its logical end, publish books in which they call free will an "illusion" and argue

for a reconstruction of society that would account for our supposed lack of agency.[20] In fact, Yet, determinism has never been proven or disproven with any level of confidence. Indeed, from the pragmatic point of view, the question of determinism or indeterminism becomes a futile quest for unattainable knowledge. Charles S. Peirce, the founder of pragmatism, refuted the idea that our belief in determinism implies the actual existence of determinism.[21] That is, we may experience reality as progressing linearly and unfolding according to prior causes, but without irrefutable evidence that no event could possibly be otherwise, the true metaphysical nature of reality is unknown. As a non-falsifiable claim, a deterministic universe does not warrant a full rebuttal in order to make the point that hard determinism is not a useful framework for the problem of implicit prejudice because it overtly denies agency.

Moving on to libertarianism, the prospects for intentionally reducing one's own implicit prejudice become more tangible. Libertarians believe that every voluntary and deliberate action an agent makes is sufficiently up to them because an agent is an uncaused cause of action. Applied to a mental process such as implicit prejudice, the framework of libertarianism would posit that agents have control over their own actions once made sufficiently aware of them. However, this argument only holds when applied to straightforward cases—such as going back to retrieve a wallet after someone points out that it was left behind—but loses coherency when applied to something as complex as implicit prejudice. Evidence from psychology research, such as the stereotype rebound effect, shows that implicit prejudice tends to evade conscious control even after someone gains awareness of it, which directly contradicts libertarian views of agency. If an agent is truly an uncaused cause, as libertarians predict, then the four methods of prejudice

[20] See: *Free Will* by Sam Harris; *Determined* by Robert Sapolsky
[21] Peirce, C. S. 1892. "The Doctrine of Necessity Examined." *The Monist* 2, (3): 321–37. http://www.jstor.org/stable/27896963.

intervention—awareness, change-based, control-based, and pharmaceutical—would likely have higher rates of success because individuals, as uncaused causes, would face little difficulty in navigating and intentionally changing their own behavior. However, as shown by the outcomes of this research, behavioral modification does not come so easily once awareness is attained. In other words, there must be something holding people back from immediate self-modification, but libertarianism does not provide the tools with which to uncover these barriers. Without an acknowledgment of the external forces that contribute to the maintenance of implicit prejudice, libertarianism overestimates the power that any one individual holds in isolation from their social environment.

Compatibilism, in finding room for agency within the constraints of determinism, solves some of the problems presented by hard determinism and libertarianism. A compatibilist free choice has more to do with desires than causes, so free will is found in actions that align with an agent's desires. However, the largest drawback of compatibilism is its tendency to adopt a binary view of agency. Agency is attributed (or denied) largely on a case-by-case basis, provided that the agent can demonstrate sufficient reasons-responsiveness in a given situation. Some situations, in particular, are seen by compatibilist thinkers as detrimental to agency. For example, a person whose compulsive disorder causes them to wash their hands every hour cannot always be said to have freely chosen to wash their hands every time. John M. Fischer delineates a standard for identifying categorically unfree behavior: "Factors that would seem to render an agent psychologically incapable of choice … might include past trauma, subliminal advertising, aversive conditioning, and even direct electric stimulation of the brain."[22] The common thread among these criteria is external influences that alter reasoning, emotional processing, and

[22] Fischer, John M., "My Compatibilism," in Pereboom, Derk (ed.), *Free Will*. Hackett Publishing Company, 2009, 250.

rational deliberation. Implicit prejudice, widely considered a form of aversive conditioning, thus constitutes a legitimate threat to free will under compatibilism. This is because implicit prejudice, as an automatic process, limits reasons-responsive mechanisms to the point of altering behavior. A person who desires to be unprejudiced, yet automatically avoids a certain group of people based solely on their skin color, is no more free—for the compatibilists—than the person who compulsively washes their hands. If freedom is defined as the ability to rationally deliberate between all possible alternatives, then implicit prejudice stands as a roadblock to rational deliberation. A compatibilist framework is highly skeptical of potential agency for people who find themselves in stifling environments because of its tendency towards a binary view of control. Arguably, a society in which racial discrimination is passed down through generations, media, and schooling would constitute such an environment. Due to the absence of mechanisms for improving control in sub-optimal environments, compatibilism fails to provide a compelling framework for understanding, and intentionally reducing, implicit prejudice.

Which concept of human agency, then, could provide an advantageous framework for constructively pursuing the problem of implicit prejudice? Such a framework should place human agency at its center while realistically acknowledging the torrent of external forces that affect the human ability to choose. I argue that John Dewey's notion of selves as systems of habits provides such a framework. Dewey's understanding of the self as embedded in material, social, and environmental systems allows for an authentic sense of agency, one which holds potential for improvement while remaining susceptible to internal and external contingencies. Understood as a system of habits, the self can be seen to enact meaningful actions within the constraints of reality because the self is an integral part of that reality. The world, according to Dewey, is a system of interdependent causes, of which humans are one. Whether causation is

deterministic is merely a tangential thread of inquiry. We can observe that action is neither fatally determined nor entirely within an individual's control by conceptualizing free will along a spectrum. The degree to which a person has agency is correlated with the extent to which they can exercise flexible habits that intelligently direct action. This pragmatic reading of agency employs a long-range conception of control that dismantles the strictly hard determinist conclusion drawn by neuroscientists and offers a more realistic concept of agency than libertarians and compatibilists. In the following section, I will provide an overview of John Dewey's pragmatic theory of human behavior and argue that a pragmatic view of free will allows for a more constructive reading of implicit prejudice and opportunities to reduce it in ourselves.

Dewey and the Self

John Dewey (1859–1952) was a prominent voice in late 19th and early 20th century American philosophy whose ideas significantly shaped modern psychology and educational theory. Dewey is widely known for his contributions to the philosophical movement of pragmatism, founded by Charles Sanders Peirce and developed by William James. He believed that knowledge and truth were not static but rather developed through human interaction and experience. His philosophical method emphasized the way in which thought is a tool for solving practical problems rather than a device for pursuing absolute truths. Dewey, furthermore, saw education as a vital means for social progress, particularly with regard to democracy, and advocated for curricula driven by experiential learning, critical thinking, and problem-solving. He rejected traditional rote learning methods in favor of an interactive, student-centered approach that encouraged curiosity. Dewey similarly emphasized the importance of participatory

democracy, arguing that an educated and engaged citizenry was essential for a functioning democratic society. His work profoundly influenced educational reform and shaped the trajectory of liberal democracy in the United States. Most importantly, for my subject matter, is how Dewey ushered in a new conception of psychology—one that broke from individualistic traditions in favor of a socially embedded, interactive theory of mind.

Much of Dewey's psychological theory was influenced by his fellow American thinker William James. James, in his *Principles of Psychology*, pioneered an experimentalist psychological theory centered on experience and observation.[23] The themes of *Principles* were advanced by Dewey, who argued that mental processes should be understood in terms of their role in helping organisms adapt to their environment. This approach contrasted with structuralist approaches which viewed mental processes as discrete elements that could be mapped onto the brain structure. The advantage of a psychology based on function, for James and later for Dewey, was a method that locates mental functions in their natural and social environments, taking up the themes of evolution and environmental fitness produced by Charles Darwin in the mid 19th century. One of James' key theories was the "reflex arc" hypothesis: an account of human behavior modeled on stimulus (cause) and response (effect).[24] James hypothesized that human behavior is a perpetual chain of cause-and-effect, where the presence of a stimulus will trigger a response, which in turn produces another stimulus, and so on. For example, a child sees a stovetop (stimulus), touches it (response), burns her hand (stimulus), and pulls her hand away (response). Breaking down an event into stimuli and responses was advantageous for identifying

[23] James, William. 1890. *The Principles of Psychology, Vol I.* The Principles of Psychology, Vol I. New York, NY, US: Henry Holt and Co. https://doi.org/10.1037/10538-000.
[24] James, *Principles*.

how individuals interact with their surroundings and created useful categories that could become testable variables in experimentation.

Dewey, however, was critical of the reflex arc hypothesis because of its tendency to isolate distinct moments of neural activation from each other, and further, to isolate the mind from the larger environment. He argued that the artificial separation of events into stimulus and response did not reflect the complex reality of behavior.[25] Imagining that organisms passively receive stimuli and then respond with an action conceals how organisms are continuously interacting with their environment; there is no concrete beginning or end to a series of actions. Furthermore, each interaction is the product of cumulative past actions and likewise modifies future actions—the child who burned her hand will be less likely to touch a hot stovetop in the future. Dewey argued that we should not strictly adhere to categories of stimulus and response because the words do not represent the true nature of a phenomenon, but rather its temporary function in a given context. A better way to view behavior, he proposed, was through the lens of habit.

The notion of habit is central to Dewey's view of human nature and likewise informs his educational and political models. In *Human Nature and Conduct*, habits are described as the general tendencies of action for an individual. Habit is "an acquired predisposition to ways or modes of response, not to particular acts," writes Dewey.[26] This distinction between modes of behavior and instances of behavior is critical because habits always have degrees of plasticity. Generally, our habits are most plastic during childhood but become rigidified through our enmeshment in social structures that encourage conformity and standardization. Not all rigidity is bad, as a certain level of behavioral stability is advantages for conserving resources. For

[25] Dewey, J. "The Reflex Arc Concept in Psychology," *Psychological Review* 3, (1896): 357-370.
[26] Dewey, *Human Nature and Conduct*, 42.

Dewey, habits are not equivalent to routines. As we constantly encounter new situations—say, a hot stovetop—we deliberate between possible modes of action. In this sense, Dewey might sound similar to a compatibilist who argues that free choice lies in the ability to choose between alternatives. But unlike the compatibilists, Dewey does not adhere to a strict affirmation of determinism and views choice as something that occurs relationally, not statically. Habits provide a cognitive template for responding to external stimuli, but they do not necessarily determine the precise way in which this response occurs. This is because habits are subject to internal and external stressors from the social and natural environment, what Dewey calls "interdependencies."[27] For example, a person may have a habit of ignoring crosswalks. The habit of jaywalking has become useful because it saves time while rushing to work in the mornings but is subject to override in the event that a car is speeding towards the intersection or a police officer is standing nearby. This example also demonstrates how habits are inextricably social: jaywalking is socially acceptable in most cities and may be customary in certain places; the person may have been encouraged to jaywalk the first time after seeing a friend do it and this action was reinforced after no one spoke out in criticism.

Changing habits also requires a transformation of underlying social conditions. Telling a person to stop jaywalking will do little to deter them, but if they were required to pay a $100 fine each time they jaywalked, this person would surely reconsider their behavior and likely change their habit. This is what Dewey means when he writes that habits must be changed "indirectly by modifying conditions, by an intelligent selecting and weighting of the objects which engage attention and which influence the fulfillment of desires."[28] To believe that habits can be changed through sheer willpower is a blatant denial of the social and environmental factors that build and

[27] Ibid., 317.
[28] Ibid., 20.

reinforce habits. It is also a misunderstanding of the way habits are formed; wanting to change is a first step, but change actually happens through extensive practice. Critical reflection and a willingness to change is often the precursor to significant behavioral modification, though habits will only take root through an intensive physical effort of reorientation. Importantly, this seemingly personal willingness, and the motivation and effort to practice, is itself the psychological artifact of a network of social causes. As environmental conditions shift, continuing to rely on habits that no longer serve our needs can perpetuate poor outcomes. Dewey calls these rigid and inflexible habits. A rigid habit is one that persists regardless of changing circumstances, preventing individuals from adjusting to new situations or engaging in critical thinking. Dewey contrasts this with flexible and intelligent habits, which allow for modification and thoughtful adjustment based on experience and context. Ultimately, every aspect of human behavior—from the desire to change, to change itself, and the tendencies that scaffold our every move—is a socially embedded system of habits.

The way that psychology conceptualizes implicit prejudice—as something hidden inside of an individual, to be exposed and extracted from the individual by means of self-awareness, conditioning, or neurochemical alterations—runs contrary to how Dewey characterizes the individual. Modern psychology, particularly within cognitive and behavioral frameworks, often treats the individual as an autonomous agent endowed with discrete internal mental processes. In *Experience and Nature*, Dewey reconceptualizes the individual not as an isolated, static entity but as an emergent, relational process defined by its ongoing interactions with its environment. Rather than viewing the individual as a preformed, self-contained subject with inherent properties, Dewey argues that what we call “an individual” is constituted through a dynamic interplay between the organism and its surroundings. He demonstrates that the privatized

characterization of the self, so common to our modern thinking, only emerged recently, while previous civilizations tended to think of the self in relation to the social groups to which it belonged. Redefining the individual is embedded in Dewey's broader critique of the subject–object dichotomy in philosophy. Instead of the individual as a subject that engages with the external world as objects, Dewey reconstructs this dichotomy to show that the individual is constituted through social relations. "[T]he mind that appears in individuals is not as such individual mind. The former is in itself a system of belief, recognitions, and ignorances, of acceptances and rejections, of expectancies and appraisals of meanings which have been instituted under the influence of custom and tradition." (219) In this view, the self is not a fixed or subjective entity but a provisional social achievement, continually shaped by and reshaped through encounters with external conditions. Thus, the individual is both a participant in and a product of an ongoing, processual engagement with nature, wherein learning, adaptation, and growth are central to the formation of identity.

Within his reconstruction of the individual, Dewey clarifies the relationship between mind and consciousness. For Dewey, mind is how we engage with the familiar, it represents a relatively stable backdrop formed by the accumulation of habits; consciousness is how we engage with the unfamiliar, when we have to deal with uncertainty that our habits cannot make immediate sense of. "Mind is contextual and persistent; consciousness is focal and transitive. Mind is, so to speak, structural, substantial; a constant background and foreground; perceptive consciousness is process, a series of heres and nows. Mind is a constant luminosity; consciousness intermittent, a series of flashes of varying intensities."[29] Dewey provides an example: when we read a book, we are immediately conscious of the meanings of words as they

[29] Dewey, John. 1925. *Experience and Nature*. Open Court Publishing Company, 303.

appear, but this momentary awareness rests upon a larger, more fundamental foundation of mind which allows us to understand language and the social meaning present in each sentence. Interestingly, there appears to be some overlap in the way that Dewey defines mind/consciousness with the way that psychologists define automatic/controlled processing. Mind, like automatic processing, seems to make up the majority of human mental activity and occurs without explicit awareness or conscious override; it is inferred to be an efficient way of making sense of the world through habit. And consciousness, like controlled processing, comes to the forefront when a person encounters scenarios that introduce doubt, which necessitates intentional and creative solution-oriented thinking. The difference, however, lies in how Dewey characterizes both mind and consciousness as two instances of the same fundamental process. As always, Dewey is trying to overcome the binary thinking that permeates philosophy, so mind and consciousness are not two separate activities but rather two aspects of the process that makes up the human self. Mind can morph into consciousness when it is challenged, and consciousness can settle into mind as it becomes a habit. Furthermore, mind and consciousness are not subjective traits belonging exclusively to an individual. They are particular instantiations of collective networks of ideas, meanings, desires, and relationships.

Flexible Habits

Dewey's reconstruction of the self becomes the foundation for his writing on freedom, agency, and causes. Implicit prejudice is a psychological phenomenon that appears to resist change on an individual level and makes intentional control very difficult. As such, it presents a twofold threat to normative conceptions of agency: firstly, it evades awareness, and secondly, it resists intentional attempts at suppression once people become aware of it. Looking at implicit

prejudice through the lenses of hard determinism, libertarianism, and compatibilism shows that this twofold threat could be interpreted pessimistically, either as an outright rejection of agency or as an instance of fatal determinism. The contemporary free will debate sees an action as "up to us" if that action originates directly from the agent, whether through will or second-order desire. Dewey's philosophical approach, on the other hand, extends the locus of control to an embodied and social conception of self, a mind that is not secluded in the inner workings of the brain. Dewey's writing can illuminate aspects of human nature that the other frameworks overlook, leading to a more fruitful discussion of implicit prejudice and its potential to be intentionally changed.

Dewey lays out two conditions for freedom that reflect the social positionality of the individual: opportunity and flexibility. The former pertains to an individual's literal ability to act, which is gained through the absence of physical and psychological restraints. In the case of implicit prejudice, opportunity comes in the form of awareness. As an automatic process, implicit prejudice tends to be hidden from awareness, which allows it to operate unseen and without conscious opposition—much like Dewey's notion of "mind," implicit prejudice serves as a backdrop to conscious reasoning. Most people do not know that they harbor implicit prejudice until something brings it to their attention, such as my taking an IAT test in class. Awareness is the first step to tackling implicit prejudice, and this is an area in which social psychology has made significant progress in awareness-based interventions. The opportunity to take an online test and receive a numerical score quantifying one's level of implicit prejudice is necessary to gain the freedom to change it. However, Dewey emphasizes that opportunity must be combined with flexibility in order to produce freedom. "But while [opportunity] is a necessary it is not a sufficient condition. *Freedom* of thought denotes freedom of *thinking*; specific doubting,

inquiring, suspense, creating and cultivating of tentative hypotheses, trials or experimentings that are unguaranteed and that involve risks of waste, loss, and error."[30] Opportunity is not sufficient for freely changing behavior because an individual's thinking—their ability to flexibly reason and intelligently deliberate—must likewise be free. Therefore, awareness is not a sufficient condition for controlling implicit prejudice. The abstract number from an IAT result means little without concrete cognitive-behavioral change, leading to the second aspect of implicit prejudice's threat to agency: awareness does not produce immediate change.

The second condition for freedom, flexibility, is Dewey's prescription for attaining improved agency. Flexibility can also be understood as the accumulation of intelligent, responsive, and creative habits. It is an individual's ability to understand their place in the world, know which behavioral opportunities exist, and flexibly choose the opportunity that best aligns with their desired outcome. "To foresee future objective alternatives and to be able by deliberation to choose one of them and thereby weigh its chances in the struggle for future existence, measures our freedom."[31] Degrees of agency, therefore, are measured by an individual's ability to purposefully modify their habits in accordance with the objective conditions in which they exist. In order to foresee alternative paths and intelligently deliberate between them requires an understanding of the interdependent forces of psychology and society, of individual and institution. Flexibility is made possible by the innate plasticity common to all human beings—a plasticity that distinguishes humanity from other, similarly evolved species. Plasticity is the ability to

However, Dewey stresses how plasticity deteriorates as people age due to the constraining nature of most social environments. As a result, flexibility, as it is derived from

[30] Ibid., 222 (emphasis in original).
[31] Dewey, *Human Nature and Conduct*, 311.

plasticity, is not enjoyed equally by everyone. People who grow up in rigid societies that enforce conformity and behavioral regularity will be less likely to express creative and flexible thinking than their open-minded, pluralistic counterparts. It is difficult to go against the social grain. "Strict repetition and recurrence decrease relatively to the novel. Apart from communication, habit-forming wears grooves; behavior is confined to channels established by prior behavior."[32] As the amount of new and stimulating information decreases in a person's lifetime, and if they tend to live according to how others live, then the habits that make up their modes of response will become solidified and difficult to change. Such is the case with implicit prejudice. But unlike compatibilists, who tend to assume that agency is dichotomous, Dewey intentionally creates room for improving agency in his philosophy, through the accumulation of degrees of flexibility.

Consciousness is the means by which individuals can interrogate and improve their flexibility, because consciousness serves as the reflective interface through which individuals become aware of, evaluate, and modify their ingrained behaviors. While mind tends to operate at an automatic, non-reflective level, consciousness provides the explicit awareness necessary to discern the inadequacies or maladaptiveness of habitual responses. When a person becomes consciously aware of their habits, they are positioned to engage in reflective inquiry—a process that entails scrutinizing the outcomes of habitual actions against one's goals or values. This evaluative process can reveal discrepancies between what is habitually enacted and what might be more beneficial or adaptive in a given context. For Dewey, the goal of flexibility is to achieve a state in which "habit is formed in view of possible future changes and does not harden so readily."[33] By acting as a catalyst for change, the intelligent use of consciousness can bring about

[32] Dewey, *Experience and* Nature, 280.
[33] Ibid., 281.

the experimental attitude that Dewey advocates: continuously testing new responses and gradually restructuring habitual patterns. Flexibility is an active, dynamic condition that underpins the possibility of intelligent habit change. It allows for the critical assessment of established behaviors, paves the way for deliberate experimentation, and bridges the gap between automatic responses and adaptive, thoughtful action.

Central to Dewey's notion of flexibility, and what sets his theory apart from traditional free will arguments, is that he locates the individual as one cause within a network of causes. Neither is the individual an uncaused cause, as with libertarianism, nor is it an unfree cog in a machine, as with hard determinism. Dewey observes that individuals are always a source of change in their environments, which is apparent from the way that situations shift in response to individual actions. "That an individual, possessed of some mode and degree of organized unity, participates in the genesis of every experienced situation, whether it be an object or an activity, is evident."[34] Thus, it is clear that individuals do have power over events, but not in the absolute way that libertarianism would suggest. It is a nuanced kind of power, subject to the multitude of social and natural forces acting alongside and through it: "There are in truth forces in man as well as without him."[35] Due largely to the way that many modern philosophies tend to isolate the individual from nature and society, most schools fail to grasp the continuity between individuals and the rest of the world. "The significance of the traditional discussion of free will is that it reflects precisely a separation of moral activity from nature and the public life of men."[36] For Dewey, the contemporary free will debate represents a misguided tendency to view morals as private affairs with no connection to social organization. As a result, philosophers of free will

[34] Ibid., 246.
[35] Dewey, *Human Nature and Conduct*, 10.
[36] Ibid., 9.

say either that (a) individuals must change themselves and that society, seen as a collection of individuals, will automatically be more moral, or that (b) society must change and individuals, as units of that society, will follow in the wake of institutional morality. Dewey sees this dichotomy as a frustrating misinterpretation of human nature, and argues that morality must proceed from both individuals and society, as the two are interdependent. His solution to this unwarranted severance comes in a recognition of the inextricable social and interdependent nature of human behavior. "[P]rogress proceeds in two ways, and that freedom is found in that kind of interaction which maintains an environment in which human desire and choice count for something."[37] Agency, and with it, morality, is neither found exclusively in the individual nor in the environment. Growth, on both the individual and collective scales, happens relationally.

The nuanced control that people have can best be thought of as indirect agency. As opposed to traditional theories that define agency as direct and immediate control over events, Dewey frames agency as something that unfolds in the course of a person's relationship with their environment. Thus, agency is not a proximate, split-second decision but an indirect, distal, and mediated series of events. "We cannot change habit directly," writes Dewey, "that notion is magic. But we can change it indirectly by modifying conditions, by an intelligent selecting and weighting of the objects which engage attention and which influence the fulfillment of desires."[38] While we, as humans, may perceive our decisions to be made in the immediate moment—since this is usually how we experience decision-making, with a tendency to place more emphasis on proximate causes—Dewey writes that what we call "a decision" is actually the culmination of a network of distal causes, of which we, as the agent, are one. In the case of implicit prejudice, this means that an individual's power to intentionally change lies not in the moment immediately

[37] Ibid., 9-10.
[38] Ibid., 20.

preceding the activation of prejudice, but in a much longer span of time leading up to moments of activation. A person may not be able to exercise control in the precise moment of prejudice activation, as evidenced by the lukewarm outcomes of control-based interventions, but they can intentionally modulate their own responses by seeking out and engaging with diverse perspectives in authentic social settings. They can alter their own habits of perception by practicing critical self-inquiry. And by improving flexibility, with creative and open-minded thinking, a person can foster habits that remain open and inviting to new, alternative possibilities of action.

III. Automaticity

In prescribing flexibility as a method for improving agency, Dewey outlines a path towards a flourishing society that fosters growth through the constant reevaluation of ideas. By resisting cultural stasis, he imagined, humanity could become reinvigorated with fresh ideas and well-functioning institutions. Evidently, this pragmatic utopia is a far cry from the current state of American society. One way to summarize the current state of affairs is a misalignment between ideals and habits. The United States has ideals of equality, meritocracy, tolerance, and pluralism. In practice, many of the habits that dictate individual and structural behavior are composed of inequality, nepotism, intolerance, and antagonism.

Dewey cautioned against naive optimism regarding cultural progress. “Old ideas do not die when the beliefs which have been explicitly associated with them disappear; they usually only change their clothes.”[39] American culture today is a clear example of old ideas hiding in modern clothing, continuing to dictate social norms under the guise of progress. In this chapter, I

[39] Dewey, *Experience and Nature*, 294.

will evaluate how prejudiced ideas precipitated in American culture, only to be swallowed by a shifting environment and become implicit prejudice. To be clear, explicit prejudice is far from absent and continues to dominate culture in many ways. My focus, instead, is to examine the way that implicit prejudice operates as an automatic habit of individuals who explicitly affirm equality, through the lens of Dewey's philosophy.

Old Ideas

Individuals perceive the world according to their habits. These habits expand beyond the boundaries of the self, as they are continuously challenged, fortified, ignored, or praised by others. Shared habits, in particular, have an incredibly strong driving force among social beings as they are constantly reinforced and strengthened. This is what Dewey calls "customs:" shared habits that propel social behavior. Customs are "active demands for certain ways of acting,"[40] which means that customs serve as moral standards for members of society. When a certain behavior is normalized, adherence is seen as morally good while deviance is seen as morally bad. In the United States, a culture of white male hegemony has historically been upheld by customs that perpetuate positive association towards the dominant culture and negative associations towards the rest of society. Three systems in particular—language, media, and schooling—have played a central role in the transmission of prejudice.

Language, for Dewey, is an extraordinary tool for creating meaning. "Of all affairs, communication is the most wonderful,"[41] writes Dewey. Language transforms natural events into social objects, with complex and dynamic meanings. Things which pass in and out of consciousness can be recalled and deliberated simply by giving them a name. Language lifts

[40] Dewey, *Human Nature and Conduct*.
[41] Dewey, *Experience and Nature*, 166.

parts of the chaotic world out of their local environment and into the broader context of human thought, giving these events a new sort of existence. “Events when once they are named lead an independent and double life.”[42] In the mind, events become laden with social meaning, they interact and challenge other mental processes; they become transformed. Meanings, far from divine truths or transcendent Platonic forms, are nothing more than the collectively-assigned, provisional labels that humans give to natural events. Learning is an immediate effect of communication, as spoken and written words are the medium by which information is transmitted across generations and between individuals. As a result, negative associations regarding specific groups of people are largely transmitted through language, even when the association is unintentional.

Linguistic transmission serves as a critical pathway through which prejudiced attitudes are communicated and internalized from a young age, constructing the social realities that shape children’s understanding of the world. As a tool for mediating experience, language scaffolds shared meanings in a society that can either disrupt or reinforce prejudice. From early childhood, the words children hear from parents, caregivers, and educators are steeped in the social habits and collective experiences of their community. Generic language, such as broad generalizations that ascribe inherent traits to entire groups (e.g., “boys are naturally assertive” or “girls are inherently nurturing”), fosters a fixed view of social identities. This framework of rigid, intrinsic qualities leads children to internalize social categories, setting the stage for prejudiced attitudes. For example, subtle linguistic variations can cue children into social hierarchies. A 2024 study by Veen, Rosanneke, and Mesman[43] found that the development of prejudice in children may be

[42] Ibid., 166.
[43] Veen, Daudi van, Rosanneke A.G. Emmen, and Judi Mesman. 2024. “Who’s to Blame? How Subtle Negative Messages about Outgroups Contribute to Ethnic Prejudice Development in Middle to Late Childhood.” *Applied Developmental Science:* 1–12. https://doi.org/10.1080/10888691.2024.2406407.

mediated by conversations between children and their mothers, depending on subtle word choices used by the mothers. Mothers who employed subtle negative language towards outgroup members were more likely to foster prejudice in their children than mothers who used neutral or positive language. Whether or not language is intentionally deployed in a prejudiced manner, the effect remains the same. In fact, Dewey points out how language continues to convey social meaning long after that meaning was originally assigned; syntax, vocabulary, and phrases become habits. "Language is always a form of action and in its instrumental use is always a means of concerted action for an end," writes Dewey.[44] This means that language is an active process involving many people. The words we speak carry an epistemic weight, as they are imbued with specific meanings that point to social objects. The "instrumental use" of language, or its purpose, is towards some end, meaning that people use language to bolster a certain viewpoint or agenda. In this way, language is political. It can reinforce the status quo or offer divergent perspectives.

Media, like language, is another way that prejudice is transmitted in society. The symbols and narratives disseminated through media do not passively convey information but actively participate in the construction of prejudice by reinforcing stereotypes. News and television programs often rely on shorthand representations to quickly convey character types, with minority groups sometimes being depicted in limited or negative roles. Such portrayals reinforce the idea of an "other" by consistently framing these groups in ways that emphasize difference and subordination, subtly shaping viewers' perceptions of social hierarchies. Additionally, advertising campaigns may idealize certain body types, skin colors, or lifestyles that align with mainstream norms, thereby reinforcing a biased view of what is considered "desirable" or

[44] Dewey, *Experience and Nature*, 184.

"normal." These portrayals do more than sell products; they contribute to a broader cultural narrative that devalues diversity and perpetuates social inequalities. Seen through Dewey's theory of the continuity between symbols and social meaning, media outputs are part of an experiential process where language, images, and narratives are intertwined, ultimately influencing how individuals perceive themselves and others. As television shows and advertisements provide recurring and widely disseminated narratives, they establish a collective social experience that normalizes prejudice in the fabric of daily life.

Schools, as epicenters of communication and learning, are one of the primary modes of transmitting customs and moral standards to children. Depending on how education is viewed, students will adopt very different habits. A school that prioritizes growth mindsets, rewards creative problem-solving, and encourages respectful debate will produce students with wide-ranging interests, resilience in the face of uncertainty, and a willingness to engage with diverse viewpoints. On the other hand, teachers who excessively punish eagerness and view students as mere recipients of knowledge will produce dull, close-minded automatons who may only succeed in tightly-controlled environments that resist change. In the latter case, students often encounter a narrow view of history and society. Textbooks may present a one-sided narrative that glorifies the achievements of a single dominant culture, while omitting or minimizing the contributions of other groups. This selective presentation of information not only distorts historical realities but also sends subtle messages about who is valued in society. For example, schools can reinforce prejudice when they present colonial history through a sanitized lens that emphasizes exploration and economic progress while omitting the brutality, exploitation, and cultural erasure that accompanied colonial expansion. When the curriculum portrays colonialism as a benign or even heroic venture, it risks normalizing systemic oppression and obscuring the

lived experiences of indigenous peoples. This imbalance in representation is a potent means by which schools transmit prejudicial attitudes, as the educational content becomes imbued with the biases of its creators and perpetuates those biases across generations. Inside of schools, media and language continue to function as transmitters of prejudice. The net result is a generation of individuals who have developed strong habits of prejudice and lack the flexible capacity to intelligently question their own beliefs.

Looking at prejudice through this lens, the distinction between implicit and explicit prejudice fades away. There is no meaningful difference between the two, as both are manifestations of deep-seated social biases that have been instilled in individuals through their participation in social environments. Prejudice often does not rise to the forefront of consciousness because people tend to maintain an egalitarian self-perception or simply lack awareness of their own motivations. In many cases, individuals may also purposefully push away signs of prejudice that bubble up in their consciousness because they do not want to identify with these negative impulses. Whether prejudice is classified as implicit or explicit is largely a function of the surrounding context—if the environment nurtures the expression of prejudice, it will become explicit, but if it suppresses overt displays, the prejudice remains hidden yet still exists beneath the surface. Ultimately, the prejudice persists because it was ingrained at some point in the individual's life through the norms and values of their community, becoming a habitual response. The distinction between implicit and explicit prejudice, while somewhat arbitrary, is useful in terms of determining how to address the issue. In cases of implicit prejudice, the first step is awareness—individuals must engage in introspection to recognize the biases that lie beneath the surface of their self-perception. Once this awareness is achieved, the challenge becomes one of changing the habitual nature of these biases through indirect control,

such as modifying the environmental cues that trigger them or adopting alternative behavioral strategies.

New Environments

What happens when the environment that once produced a set of customs suddenly changes? Instead of changing alongside the environment, customs remain steadily in place, even when they do not reflect the new social circumstances. This is because customs, as shared habits, are susceptible to the same conservative rigidity that affects individuals. The conservation of prejudiced customs can be analogized to Dewey's discussion of myths:

> In one situation fancy generates stories which are consistent with desire and are attractive. These are connected with ceremonies to which, in addition to their immediate good, external efficiency is imputed. They become nuclei about which observations and ideas continually gather; they are centres of mental as well as emotional systematizations. It is no wonder that myths long prevail. When the development of industry and tested inquiry makes it evident that the actual world will not accept them nor stand for them, their actuating springs remain in full force and the river of revery still flows.[45]

Each of these statements regarding myths can be mapped onto prejudice. In the first place, prejudice arises ("generates stories") in order to fit a sociopolitical narrative, such as white supremacy. The myths that were popularized by scientific racism in the 18th and 19th centuries, such as phrenology and false theories of evolution, were used to justify the inherent superiority of white Europeans and naturalize colonial enslavement. These falsehoods were accompanied by cultural rituals ("connected with ceremonies"), such as public slave auctions, minstrel shows,

[45] Ibid., 226-7.

and Confederate pageantry. The cultural narrative of white supremacy became so embedded in the foundation of American statehood that it served as a moral framework. This is what Dewey means when he writes that myths "become nuclei" and "centers of … emotional systematizations." As a person grows up in this society, each new experience is held up against the backdrop of a preexisting ideology, which is taken to be an objective truth. Without the discernment afforded by flexibility, most people tend to disregard experiences that do not match their preconceived notions of reality.

However, as Dewey notes, circumstances do change. While racism, sexism, and other forms of prejudice are still widely prevalent, there is a greater expectation of equal treatment today. Since the Civil Rights Movement, America has made significant strides toward legal equality, enshrining rights and protections intended to bolster equal treatment. Legislation such as the Civil Rights Act of 1964 and the Voting Rights Act of 1965 marked transformative milestones by prohibiting discrimination based on race, color, religion, sex, or national origin and ensuring broader electoral participation. These legal changes have altered the country's constitutional framework, making it clear that all citizens are entitled to equal treatment under the law. On paper, the nation now boasts a legal system that, in theory, guarantees equal access to opportunities in education, employment, housing, and public accommodations, reflecting an evolution toward a more inclusive society. The second part of Dewey's quote about myths alludes to shifting social expectations, and he points out how oftentimes society will reevaluate moral standards in light of new evidence, thus explicitly disavowing the myths that characterized an earlier time. However, the central claim of Dewey's argument is that habits of thought and action will continue to operate long after society has rejected them. "Political and legal institutions may be altered, even abolished; but the bulk of popular thought which has been

shaped to their pattern persists,"[46] writes Dewey, referring to the tendency of rigid, unreflective habits to persist in any environment. Such persistence of popular thought is not merely inertia but the residue of past experiences that continue to shape present ideas. Institutions, whether political or legal, function as frameworks for communal experience, and the values and norms they propagate are internalized over time. The source of a habit remains at full strength even as external variables change around it.

To illustrate this point, several critical race theorists have pointed out how the overt language of white supremacy has been substituted for subliminal phrases meant to convey the same meaning. Patricia Hill Collins, for example, writes: "Racism didn't magically go away just because we refuse to talk about it. Rather, overt racial language is replaced by covert racial euphemisms that reference the same phenomena."[47] These scholars challenge the notion that racism died with the Civil Rights Movement, instead arguing that it has evolved into a form that can adapt to a new environment. One need not adopt the methodology of critical race theory, nor agree with all of its conclusions, to affirm the validity of these sentiments. It is evident that Dewey would agree: "A habit impeded in overt operation continues nonetheless to operate."[48] In this case, the use of racial slurs has been impeded in overt operation by new moral standards, however, the same sentiments continue to manifest as not-so-subtle disparaging comments. This process explains why implicit prejudice operates automatically. The habits of racial and gendered inequality that shape a person's perception of, and interaction with, outgroup members continue to exist long after it is no longer socially acceptable to be racist or sexist. The modes of communication that transmit stereotypes, such as language, media, and education, continue to

---

[46] Dewey, *Human Nature and Conduct*, 108.
[47] Collins, Patricia Hill. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. Boston: Unwin Hyman.
[48] Dewey, *Human Nature and Conduct*, 53.

operate in subtle but impactful ways. This is why people who gain awareness of their own implicit prejudice through an IAT exam do not yet have the sufficient means to change it: the habits that underlie prejudice, and the social institutions that scaffold it, are continuing to operate beneath the surface of awareness, pushing back on an individual's attempt to become unprejudiced. If the self is conceptualized only as a subjective, inner mind that is shielded from the outer world, then change would be straightforward and immediate. But since, as Dewey articulates, the self is a process by which individuals engage with their environment, change must come from the environment as much as from the individual.

IV. Improving Flexibility

So far, it has been established that individuals can exercise agency to dismantle their own implicit prejudice, where agency is the capacity to improve one's own degree of flexible habits. Automaticity has been reconstructed as the stubborn habits of prejudice that remain operative in a new environment, so change must address the source of these habits. Current attempts at reducing implicit prejudice center around the task of turning implicitly prejudiced people into unprejudiced people. Making people aware of their own implicit prejudice, instructing them to change their explicit associations, attempting to bolster control at the moment of prejudice activation, and even going so far as to administer drugs that can reduce prejudice—each of these interventions is clearly aimed at redirecting behavior from one modality (implicit prejudice) into another modality (unprejudiced).

The goal of interventions must shift. Instead of seeking to make people less prejudiced, the emphasis should be placed on helping people improve their flexibility. At first, this might appear to be a counterintuitive strategy—shouldn't we be fighting prejudice directly? However,

as I will show, orienting towards flexibility has greater potential to disrupt habits of prejudice because it more realistically accounts for the complexity of human behavior.

A Pathology of Goodness

Dewey issues a warning to those who believe that goodness can be found in the separation of individuals from their environment: "There is a pathology of goodness as well as of evil."[49] This kind of goodness is only "good" in the sense that it furthers an agenda based on an incorrect understanding of human nature. What this means, in practice, is that people will strive to conform to rigid societal expectations of goodness, by seeking praise and avoiding punishment, instead of thinking critically about which action to take. Nowhere is this fact more exemplified than with the phenomenon of benevolent prejudice.

Benevolent prejudice is a form of implicit prejudice that refers to evaluations of outgroups that appear positive on their surface while perpetuating negative outcomes for the objects of those evaluations. It is most commonly identified with beliefs about women and so-called "model minorities." The defining feature of benevolent prejudice is how it selectively identifies traits and behaviors that, when praised, work to uphold the system of inequality that positions outgroup members as inferior to oneself. For example, benevolent sexism might describe women as caring and nurturing; "pure creatures who ought to be protected, supported, and adored."[50] The practical result of this characterization is that women are confined to domesticity and subjected to patriarchal domination in the private and public spheres. Similarly, Asian Americans are subject to the "model minority" caricature which posits them as

---

[49] Ibid., 4.

[50] Glick, P., & Fiske, S. T. 2001. An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56(2), 109–118. https://doi.org/10.1037/0003-066X.56.2.109

hardworking and academically competent. Data from the Pew Research Center in 2023 shows that nearly two-thirds of Asian American adults experience these stereotypes in day-to-day encounters.[51] A likely outcome is the exclusion of Asian Americans from leadership roles; the NIH notes that while Asian Americans are overrepresented in the healthcare labor market, they receive only 6% of leadership positions.[52]

As a form of implicit prejudice, benevolent prejudice is especially insidious in its attempts to hide behind its own facade of positivity. To express seemingly positive views of women or racial minorities is more acceptable in our society than outright hostility towards outgroups. The superficial positivity of benevolent prejudice places its targets in awkward and uncomfortable positions when they try to defend themselves, especially because the context of benevolent prejudice is often subtle and invisible to outside observers. The existence of benevolent prejudice pushes back against the notion that prejudice must exclusively constitute a negative view of others. It demonstrates how the impact of prejudice can be harmful even when the prejudice takes on a sympathetic or lighthearted form, as it can be used to justify egregious discrimination and oppression. The goal of interventions, therefore, must be the cultivation of reflective and purposeful egalitarian agents. To set our sights merely on the elimination of harmful stereotypes and implicit hostility will result in people who avoid blame without changing their underlying habits.

---

[51] Ruiz, N.I, et al. 2023. “Asian Americans and the ‘Model Minority’ Stereotype.” Discrimination Experiences Shape Most Asian Americans’ Lives. *Pew Research Center.* https://www.pewresearch.org/race-and-ethnicity/2023/11/30/asian-americans-and-the-model-minority-stereotype/.

[52] Goon, C., et al. 2022. “Examining the Asian American leadership gap and inclusion issues with federal employee data: Recommendations for inclusive workforce analytic practices.” *Frontiers in Research Metrics and Analytics*, 7, 958750. https://doi.org/10.3389/frma.2022.958750

Adaptive Agents

What benevolent prejudice demonstrates is that surface-level positivity is not the same thing as being unprejudiced. This could be part of the reason why current interventions do not work as well as intended, because participants may be masking their implicit prejudice with benevolence instead of critically addressing their own hostility. It is for this reason that I argue interventions must reorient towards the development of flexible, adaptive agents.

Dewey defines flexibility as the capacity of individuals to give equal weight to all behavioral possibilities before intelligently choosing a path of action. In this way, Dewey is calling for a clear redress to the way that prejudice operates—recalling how *prejudice* refers to a pre-judgment; the tendency of individuals to prioritize certain stimuli more than others, and to act in pursuit of a certain end without thoughtfully considering each opportunity. The solution lies not in pursuing different ends (that of being unprejudiced) but, rather, of minimizing the pursuit of ends entirely. For becoming absorbed with any end in particular is, for Dewey, the antithesis of flexibility. Environments continuously shift, social institutions rise and fall, and the flexible agent is one who can creatively evaluate choices without clinging to any outcome or resolution.

> Deliberation is irrational in the degree in which an end is so fixed, a passion or interest so absorbing, that the foresight of consequences is warped to include only what furthers execution of its predetermined bias. Deliberation is rational in the degree in which forethought flexibly remakes old aims and habits, institutes perception and love of new ends and acts.[53]

[53] Dewey, *Human Nature and Conduct*, 198.

Rational deliberation is the key to improving agency, and improving agency is the key to overcoming implicit prejudice in oneself. While creating positive associations of outgroup members can be useful in introducing people to new habits of perceptions, these habits will not form lasting impressions if they are framed under the guise of sheer positivity. This is evidenced by the rebound effect, which shows people putting up psychological resistance to counter-stereotype training, as well as the prevalence of benevolent prejudice. A better alternative is to guide people towards flexible habits, where they can become more responsive to a wider range of possibilities.

A determinist might, at this point, raise an objection. Isn't this just another form of manipulating people into the kind of person that you want them to be? How is this a form of agency? In fact, there are philosophers who use the position of hard determinism to argue against the possibility of agency in counteracting prejudice. David Wellman takes the neurological fact of intergroup prejudice to be an argument for the "inescapability" of racial prejudice, insofar as we understand prejudice to be a psychological mechanism. Wellman uses this line of thinking to argue for an understanding of racial prejudice as a sociological mechanism, positing that we should disregard the neural explanation altogether.

> If bias is ultimately a function of biology and neurology, human actors do not control it. Consequently, they cannot be held accountable for discriminatory behavior. The human actor in this account is a medium through which the normal cognitive processes of categorization are played out. They are passive recipients of raw data, which the brain then processes into categories that automatically bias what they see and how they act.[54]

[54] Wellman, David. 2007. "Unconscious Racism, Social Cognition Theory, and the Legal Intent Doctrine: The Neuron Fires Next Time." *Handbooks of the Sociology of Racial and Ethnic Relations*, 50.

Here, Wellman paints a picture of racial prejudice as an inevitable firing of neurons that determines human behavior. Viewing racial prejudice this way would preclude the kind of change that originates in an agent: "There is no space for intervention or change, except for undefined, Orwellian notions like 'mental correction' or 'careful process re-engineering.'"[55]

In response, I would point out how the goal of creating a flexible agent is to foster an individual who can consider alternative possibilities. This inherently allows for disagreement. In considering a range of potential actions, a flexible agent will most certainly consider actions that reproduce prejudice. They will consider greater possibilities than would be realistic from a place of rigidity. The point, in improving flexibility, is that an agent who seeks to dismantle prejudice will disregard those actions if they find them to be harmful. The commitment is to a process, not to specific outcomes. A useful way to think about flexibility is the following: instead of getting people to *do a thing* (i.e. automatons who follow a script), flexibility encourages people to become *the kind of person who does that thing* (i.e. an adaptive agent). Ultimately, the person is still free to do what they want, yet, they will have developed habits and dispositions that orient them towards genuine equality. Instead of pinpointing one belief that must be changed, the entire self is reoriented towards egalitarianism.

Conclusion

In this thesis, I asked whether people can intentionally change their own implicit prejudice. As an automatic process that appears to resist both awareness and intentional change, implicit prejudice constitutes a clear threat to agency under many of the leading theories of free will. Yet, by examining implicit prejudice through the lens of John Dewey's philosophy—of

[55] Wellman, "Unconscious Racism," 51.

humans as systems of habits, and of agency as degrees of flexible habits—I defended the position that agents do have indirect control over their own implicit prejudice. This kind of control is neither immediate nor absolute. It requires a profound critical reflection into one's own psyche, as well as the substantive work of recognizing and changing one's own habits. But it is possible.

In Part I, I reviewed the way that psychologists think about implicit prejudice in order to lay the groundwork for my subsequent analysis. I identified the key terminology associated with implicit prejudice research, and traced its development from conceptual beginnings to present status as a cultural buzzword. I also reviewed the four main ways that psychologists are attempting to help people combat their own implicit prejudice: awareness, change-based, control-based, and pharmaceutical interventions. While each method has beneficial aspects, none of the four interventions currently achieve the lasting cognitive-behavioral modifications that they try to achieve. This led me to my guiding question: is it possible for people to change their own implicit prejudices, and if so, how?

In order to answer this question, in Part II, I looked for a theory of agency that asserts the possibility of control while remaining realistic about the difficulties posed by implicit prejudice. Starting with traditional theories of hard determinism, libertarianism, and compatibilism, I applied each of their frameworks to the problem of intentionally reducing one's own implicit prejudice. Hard determinism was a nonstarter, due to its denial of agency; libertarianism was not useful, since it treats awareness as a sufficient condition for control; and compatibilism did not work either, since it would likely classify implicit prejudice as a fatal threat to acting in a way that aligns with one's desires. Instead, I turned to John Dewey's theory of humans as systems of habits as an alternative conception of agency. By reconstructing agency as degrees of flexible

habits, and recognizing the continuity between individuals and social environments, Dewey shows that people can gain control in areas where it has been stunted by rigid habits. Dewey emphasizes practice as well as ideas, positing that freedom is gained through the accumulation of flexibility and awareness.

In Part III, I used Dewey's writing on customs and education to closely analyze the way that implicit prejudice is constructed and maintained in society, and persists even when that society switches course towards overt equality. What psychologists call an "automatic process" can be understood, in Deweyan terms, as old ideas stubbornly resisting new environments. Institutions of language, media, and schooling work to perpetuate habits of prejudice, creating a phenomenon where people continue to hold prejudice even when they explicitly enforce equality. By acknowledging how individuals exist within a network of causes, a clearer picture of implicit prejudice can be drawn.

In Part IV, I argued that interventions, which usually aim to make people unprejudiced by directing them towards a certain end, may become more effective if they instead aim to produce flexible agents. One of the problems with current interventions is that they emphasize a rigid view of goodness which—while not bad in specific social environments—may inevitably become altered over the course of time. These interventions already show signs of ineffectiveness, such as the issue of benevolent prejudice in place of genuine habit change. A better way of looking at the problem of implicit prejudice is instead to foster the growth of people who can intelligently deliberate in any environment. The improvement of flexible deliberation will likely lead to agents who do not act on prejudice because they are more able to critically evaluate their own preconceptions.

Dewey's pragmatism is melioristic, meaning that it is neither optimistic nor pessimistic about the outcome of a given process. It does say, however, that by working intelligently towards a goal and remaining flexible in one's approach to a solution, we may set ourselves up for a more successful outcome. In his reconstruction of the human psyche and what it entails for agency, morality, and society, Dewey never claims to have the answers to any of the world's problems. Instead, he offers his readers thoughtful hypotheses about the human condition, based on the best scientific evidence available to him at the time, in order to orient us towards progress.

> A morals based on study of human nature … would not automatically solve moral problems, nor resolve perplexities. But it would enable us to state problems in such forms that action could be courageously and intelligently directed to their solution. It would not assure us against failure, but it would render failure a source of instruction.[56]

Dewey wrote in search of a philosophy that could learn from its own mistakes. His goal was not to solve the problem of morality in one fell swoop—as he knew that any momentary solution must inevitably change to accommodate a shifting environment—but rather to locate weak points in social theory and effectively apply resources towards their improvement, allowing society to build a wealth of knowledge from which progress can be steadily attained. And in doing so, he regarded the study of human nature as one of the most vital sources of instruction. Habits are the building blocks of human nature, the foundation on which Dewey's social and political theories are built. This is because human nature is both social and political. As habits of exclusion, dominance, and social hierarchy are formed, they are created and reinforced through social and political structures. It is those same social and political structures, the interplay of psychology

[56] Dewey, *Human Nature and Conduct*, 12.

with its environment, that must change if society is to cultivate flexible agents who can work to reduce their own implicit prejudice.

Arguing for something in theory does not mean that it will automatically come into practical fruition. Just like implicit prejudice requires awareness *and* practice in order to be intentionally reduced, the problem itself will not be solved through the writing of an undergraduate thesis paper. Writing merely orients us in the direction we need to go. Recognizing the possibility of intentional change is only a first step in the long road to enacting real social change. Insights from this paper, on the formation and habit-driven reproduction of implicit prejudice, must be applied to real-world situations, with real people. Community organizing, education reform, government, policy-making, and journalism are all potential vessels for the practical enactment of this work. In a pluralistic, diverse society like the United States, it is often up to local communities to identify which programs will work best in their unique settings. The important aspect is not the "what" but the "how:" an effective strategy will locate individuals within their proximate and distant communities, model positive behavior, encourage the formation of intelligent habits, foster respectful disagreement and lively debate, and hold people accountable for long-term growth. It will foster the development of people who are intelligent and creative problem-solvers, not merely primed to a certain behavior, but empowered to visualize a vast array of possibilities and deliberate between them, choosing the paths that lead to the most good.

Bibliography

"A History of Implicit Social Cognition: Where Is It Coming from? Where Is It Now? Where Is It Going?" n.d. Accessed March 12, 2025. https://psycnet.apa.org/record/2010-13147-001.

"About the IAT." n.d. Accessed March 12, 2025. https://implicit.harvard.edu/implicit/iatdetails.html.

Collins, Patricia Hill. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment.* Boston: Unwin Hyman.

Dewey, John. 1896. "The Reflex Arc Concept in Psychology." *Psychological Review* 3:357–70.

Dewey, John. 1922. *Human Nature and Conduct: An Introduction to Social Psychology*. Henry Holt And Company.

Dewey, John. 1925. *Experience and Nature*. Open Court Publishing Company.

Fischer, J. M. 2013. "My Compatibilism." In (eds. Paul Russell and Oisin Deery), *The Philosophy of Free Will: Essential Readings From the Contemporary Debates*. Oxford University Press.

Glick, P., & Fiske, S. T. 2001. An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, 56(2), 109–118. https://doi.org/10.1037/0003-066X.56.2.109

Goon, C., et al. 2022. "Examining the Asian American leadership gap and inclusion issues with federal employee data: Recommendations for inclusive workforce analytic practices." *Frontiers in Research Metrics and Analytics*, 7, 958750. https://doi.org/10.3389/frma.2022.958750

Greenwald, A. G., and M. R. Banaji. 1995. “Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes.” *Psychological Review* 102 (1): 4–27. https://doi.org/10.1037/0033-295x.102.1.4.

Harris, Sam. 2012. *Free Will*. Free Press.

“Home | Emory University | Atlanta GA.” n.d. Accessed March 12, 2025. https://diversity.emory.edu/index.html.

Howell, Jennifer L., Sarah E. Gaither, and Kate A. Ratliff. 2015. “Caught in the Middle: Defensive Responses to IAT Feedback Among Whites, Blacks, and Biracial Black/Whites.” *Social Psychological and Personality Science* 6 (4): 373–81. https://doi.org/10.1177/1948550614561127.

“Implicit Bias.” n.d. Accessed March 12, 2025. https://www.apa.org/topics/implicit-bias.

“Implicit Bias | Change-Based Interventions.” *Stanford Encyclopedia of Philosophy.* Accessed March 12, 2025. https://plato.stanford.edu/entries/implicit-bias/#ChanBaseInte.

“Implicit Bias | Control-Based Interventions.” *Stanford Encyclopedia of Philosophy.* Accessed March 12, 2025. https://plato.stanford.edu/entries/implicit-bias/#ContBaseInte.

“Inclusive Pedagogy | Emory University | Atlanta GA.” n.d. Accessed March 12, 2025. https://cfde.emory.edu/programs/teaching/inclusive-pedagogy/index.html.

James, William. 1890. *The Principles of Psychology, Vol I.* The Principles of Psychology, Vol I. New York, NY, US: Henry Holt and Co. https://doi.org/10.1037/10538-000.

Kahneman, D. 2011. *Thinking, Fast and Slow*. Macmillan.

Kite, Mary E., Bernard E. Whitley, and Lisa S. Wagner. 2022. “Psychology of Prejudice and Discrimination.” Routledge & CRC Press. July 25, 2022.

https://www.routledge.com/Psychology-of-Prejudice-and-Discrimination/Kite-WhitleyJr-Wagner/p/book/9780367408176.

"Long-Term Reduction in Implicit Race Bias: A Prejudice Habit-Breaking Intervention - PMC." Accessed March 12, 2025. https://pmc.ncbi.nlm.nih.gov/articles/PMC3603687/.

Madva, A. 2016. "A Plea for Anti-Anti-Individualism: How Oversimple Psychology Misleads Social Policy." *Ergo: An Open Access Journal of Philosophy* (3): 701-728.

Nosek B. A., Greenwald A. G., Banaji M. R. 2005. Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Psychological Bulletin*, 31(2):166-80. DOI: 10.1177/0146167204271418.

Payne, B. K., & Gawronski, B. 2010. A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*, 1-15. The Guilford Press.

Peirce, C. S. 1892. "The Doctrine of Necessity Examined." *The Monist* 2, (3): 321–37. http://www.jstor.org/stable/27896963.

"President Obama's Farewell Address." n.d. The White House. Accessed March 12, 2025. https://obamawhitehouse.archives.gov/node/360231.

Ruiz, N., et al. 2023. "Asian Americans and the 'Model Minority' Stereotype." Discrimination Experiences Shape Most Asian Americans' Lives. *Pew Research Center*. https://www.pewresearch.org/race-and-ethnicity/2023/11/30/asian-americans-and-the-model-minority-stereotype/.

Sapolsky, R. 2023. *Determined: A Science of Life Without Free Will*. Penguin Press.

Shen, Z., Hunt, J. 2008. The stereotype rebound effect: Universal or culturally bounded process?,

Journal of Experimental Social Psychology (44): 3, 489-500. https://doi.org/10.1016/j.jesp.2007.07.010.

Terbeck, S., Kahane, G., McTavish, S. et al. 2012. Propranolol reduces implicit negative racial bias. *Psychopharmacology* (222), 419–424. https://doi.org/10.1007/s00213-012-2657-5

Wellman, David. 2007. "Unconscious Racism, Social Cognition Theory, and the Legal Intent Doctrine: The Neuron Fires Next Time." *Handbooks of the Sociology of Racial and Ethnic Relations*: 39-65.