IMPLICIT WHITE FAVORITISM IN THE CRIMINAL JUSTICE SYSTEM

Robert J. Smith,[*] *Justin D. Levinson,*[**] *and Zoë Robinson*[***]

ABSTRACT

Commentators idealize a racially fair criminal justice system as one without racial derogation. But unjustified racial disparities would persist even if racial derogation disappeared overnight. In this Article, we introduce the concept of implicit white favoritism into criminal law and procedure scholarship, and explain why preferential treatment of white Americans helps drive the stark disparities that define America's criminal justice system. Scholarly efforts thus far have shone considerable light on how implicit negative stereotyping of black Americans as hostile, violent, and prone to criminality occurs at critical points in the criminal justice process. We rotate the flashlight to reveal implicit favoritism, a rich and diverse set of automatic associations of positive stereotypes and attitudes with white Americans. White favoritism can operate in a range of powerful ways that can be distinguished from traditional race-focused examples: in the way, for example, white drivers are pulled over less often than unseen drivers or crimes against white victims are seen as more aggravated. Our account of implicit white favoritism both enriches existing accounts of how implicit racial bias corrupts the criminal justice system and provides explanations for disparities that implicit negative stereotyping explanations miss altogether.

INTRODUCTION

Commentators idealize a racially fair criminal justice system as one without racial derogation. But unjustified racial disparities would persist even if derogation disappeared overnight.[1] In this Article, we introduce the

* Assistant Professor of Law, University of North Carolina School of Law.

** Professor of Law, William S. Richardson School of Law, University of Hawai'i.

*** Associate Professor of Law, DePaul University College of Law.

1. Racial disparities pervade the American criminal justice system. Commentators almost universally agree on this fact. *See, e.g.*, MICHELLE ALEXANDER, THE NEW JIM CROW: MASS INCARCERATION IN THE AGE OF COLORBLINDNESS 16 (rev. ed. 2012) ("The fact that more than half of the young black men in many large American cities are currently under the control of the criminal justice system (or saddled with criminal records) is not—as many argue—just a symptom of poverty or poor choices, but rather evidence of a new racial caste system at work."); Ian F. Haney López, *Post-Racial Racism: Racial Stratification and Mass Incarceration in the Age of Obama*, 98 CALIF. L. REV. 1023, 1028 (2010) ("Even the most cursory engagement with American criminal justice at the start of

concept of implicit white favoritism into criminal law and procedure scholarship, and explain why preferential treatment of white Americans helps drive the stark disparities that define America's criminal justice system.[2]

Our project is part of the larger scholarly effort over the past decade to draw on an extensive body of social science that demonstrates how individual actors in the criminal justice system—and in society generally—possess implicit racial biases that can affect their perceptions, judgments, and behaviors.[3] Criminal law scholars have employed implicit bias-based analyses to help explain racial discrepancies in police stop-and-frisk rates,[4] arrest rates,[5] prosecutorial charging and bargaining,[6] sentencing,[7] and other

---

the twenty-first century drives home the twin points that the United States puts people under the control of the correctional system at an anomalously high rate, and that it shuts behind bars an overwhelmingly disproportionate number of black and brown persons.").

2. Our focus on implicit white favoritism has intellectual roots in the literature on white privilege. For the classic treatment, see generally Peggy McIntosh, *White Privilege: Unpacking the Invisible Knapsack*, *in* RACE, CLASS, AND GENDER IN THE UNITED STATES 188, 198–90 (Paula S. Rothenberg ed., 6th ed. 2004) (cataloguing "some of the daily effects of white privilege" and listing, for example, "[i]f a traffic cop pulls me over or if the IRS audits my tax return, I can be sure I haven't been singled out because of my race"). For a more recent treatment, see generally Stephanie M. Wildman, *The Persistence of White Privilege*, 18 WASH. U. J.L. & POL'Y 245 (2005) (explaining white privilege, in part, as a treatment of whiteness as neutral—a control variable of sorts—and considering how such privilege is perpetrated). Scholars commonly use the term "white privilege" to mean those benefits that flow to white Americans by virtue of the absence of racial derogation directed at them. *See, e.g.*, Devon W. Carbado, *(E)Racing the Fourth Amendment*, 100 MICH. L. REV. 946, 1002 (2002) ("If we assume, for example, that nonwhite people are more vulnerable to being stopped by the police than whites, more likely than whites to feel apprehensive about such encounters, and less likely than whites to know and feel empowered to exercise their constitutional rights, ignoring this racial disparity privileges whites. It prefers their racial position vis-à-vis police interactions to the racial position of nonwhites."). In this Article, we use "white favoritism" in a very particular way: to connote those benefits conferred to white Americans because of their whiteness *relative to a baseline of race neutrality*. In reality, derogation and favoritism work in tandem to "privilege whites," but favoritism is an independent mechanism that necessitates separate treatment.

3. On implicit bias in the law generally, see IMPLICIT RACIAL BIAS ACROSS THE LAW (Justin D. Levinson & Robert J. Smith eds., 2012). *See also* Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CALIF. L. REV. 945, 966 (2006) ("[I]mplicit race bias is pervasive and is associated with discrimination against African Americans."); Jerry Kang & Kristin Lane, *Seeing Through Colorblindness: Implicit Bias and the Law*, 58 UCLA L. REV. 465 (2010); Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489 (2005); Justin D. Levinson, *Forgotten Racial Equality: Implicit Bias, Decisionmaking, and Misremembering*, 57 DUKE L.J. 345, 354 (2007); Justin D. Levinson et al., *Guilt by Implicit Racial Bias: The Guilty/Not Guilty Implicit Association Test*, 8 OHIO ST. J. CRIM. L. 187 (2010); Janice Nadler, *No Need to Shout: Bus Sweeps and the Psychology of Coercion*, 2002 SUP. CT. REV. 153 (claiming that the Supreme Court's consent jurisprudence should include findings from the psychology of compliance).

4. L. Song Richardson, *Cognitive Bias, Police Character, and the Fourth Amendment*, 44 ARIZ. ST. L.J. 267 (2012); L. Song Richardson, *Police Efficiency and the Fourth Amendment*, 87 IND. L.J. 1143, 1145 (2012) (arguing that "[i]mplicit social cognition research demonstrates that implicit biases can affect whether police interpret an individual's ambiguous behaviors as suspicious"); Josephine Ross, *Can Social Science Defeat a Legal Fiction? Challenging Unlawful Stops Under the Fourth Amendment*, 18 WASH. & LEE J. CIVIL RTS. & SOC. JUST. 315 (2012).

5. Shima Baradaran, *Race, Prediction, and Discretion*, 81 GEO. WASH. L. REV. 157, 164–65 (2013) (examining police behavior and implicit bias specifically in the context of drug arrests); L. Song

areas where disparities persist.[8] These scholars have demonstrated, project by project, that implicit negative stereotypes of black Americans pervade the American psyche. For example, Americans rate ambiguous pieces of evidence to be more probative of guilt when a suspect is dark-skinned and display a stronger implicit connection between "black" and the concept "guilty" than they do between "white" and "guilty."[9] The overriding theme in this work is that implicit negative stereotypes of black Americans as hostile, violent, and prone to criminality create a lens through which criminal justice actors automatically perpetuate inequality.

This picture is incomplete. Even if we could eliminate the bias that these scholars have illuminated, racial disparities would persist because removing derogation is not the same as being race-neutral. If legislators, police officers, jurors, and legal professionals implicitly favor white Americans then we still possess a racialized justice system. To gain a fuller understanding of what drives unjustified disparities, then, we must rotate the flashlight ever so slightly to reveal a rich and diverse form of implicit racial bias that has been overlooked in criminal law and procedure research. This is the bias of implicit favoritism. Implicit favoritism can be defined as the automatic association of positive stereotypes and attitudes

---

Richardson, *Arrest Efficiency and the Fourth Amendment*, 95 MINN. L. REV. 2035, 2036 (2011) (claiming that "implicit social cognition . . . can contribute much to the understanding of police behavior"); *see also* Mary Nicol Bowman, *Full Disclosure: Cognitive Science, Informants, and Search Warrant Scrutiny*, 47 AKRON L. REV. 431 (2014); Robert D. Crutchfield, *Warranted Disparity? Questioning the Justification of Racial Disparity in Criminal Justice Processing*, 36 COLUM. HUM. RTS. L. REV. 15, 20 (2004) (examining the extent to which racial differentials in arrests actually represent crime involvement given the potential bias in some jurisdictions).

6. Robert J. Smith & Justin D. Levinson, *The Impact of Implicit Racial Bias on the Exercise of Prosecutorial Discretion*, 35 SEATTLE U. L. REV. 795 (2012).

7. *See, e.g.*, Jerry Kang et al., *Implicit Bias in the Courtroom*, 59 UCLA L. REV. 1124 (2012); Justin D. Levinson et al., *Devaluing Death: An Empirical Study of Implicit Racial Bias on Jury-Eligible Citizens in Six Death Penalty States*, 89 N.Y.U. L. REV. 513 (2014); Robert J. Smith & G. Ben Cohen, *Capital Punishment: Choosing Life or Death (Implicitly)*, *in* IMPLICIT RACIAL BIAS ACROSS THE LAW 229 (Justin D. Levinson & Robert J. Smith eds., 2012).

8. *See, e.g.*, Stephen Demuth, *Racial and Ethnic Differences in Pretrial Release Decisions and Outcomes: A Comparison of Hispanic, Black, and White Felony Arrestees*, 41 CRIMINOLOGY 873, 898 (2003) (finding that Hispanic defendants are more likely to be detained than white and black defendants, and racial/ethnic differences are most pronounced in drug cases); Marvin D. Free, Jr., *Racial Bias and the American Criminal Justice System: Race and Presentencing Revisited*, 10 CRITICAL CRIMINOLOGY 195, 220 (2001) (finding evidence of discrimination in key criminal justice decision points); Andrea D. Lyon, *Race Bias and the Importance of Consciousness for Criminal Defense Attorneys*, 35 SEATTLE U. L. REV. 755 (2012); L. Song Richardson & Philip Atiba Goff, *Implicit Racial Bias in Public Defender Triage*, 122 YALE L.J. 2626 (2013) (examining implicit racial bias in the context of the public defender's office); Cassia Spohn, *Race, Sex, and Pretrial Detention in Federal Court: Indirect Effects and Cumulative Disadvantage*, 57 U. KAN. L. REV. 879, 898–99 (2009) (finding that being under the control of the criminal justice system increased the odds of pretrial detention for blacks but not for whites).

9. *See* Justin D. Levinson & Danielle Young, *Different Shades of Bias: Skin Tone, Implicit Racial Bias, and Judgments of Ambiguous Evidence*, 112 W. VA. L. REV. 307 (2010); Levinson et. al., *Guilt by Implicit Bias*, *supra* note 3.

with members of a favored group, leading to preferential treatment for persons of that group.[10] In the context of the American criminal justice system, implicit favoritism is white favoritism.[11] Consider a white police officer deciding whether to stop a vehicle. The favoritism question is not whether the officer is less likely to pull over a white driver than a black driver, but rather if the officer is less likely to pull over a vehicle when she sees a white driver than when she is entirely unaware of the driver's race.

In this Article, we explain that implicit favoritism is important because it helps to drive racial disparities in the criminal justice system. Social scientists have linked implicit favoritism to the ability of jurors to accurately remember damning details of an alleged offense,[12] to the evaluation of whether negative actions taken by another are the result of one's disposition or instead to the circumstances that constrained one's choices,[13] and to the degree of empathic response to human pain.[14] Implicit

10. *See e.g.*, Nilanjana Dasgupta, *Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations*, 17 SOC. JUST. RES. 143 (2004); Nilanjana Dasgupta, *Ingroup Experts and Peers as Social Vaccines Who Inoculate the Self-Concept: The Stereotype Inoculation Model*, 22 PSYCHOL. INQUIRY 231 (2011); Anthony G. Greenwald & Thomas F. Pettigrew, *With Malice Toward None and Charity for Some: Ingroup Favoritism Enables Discrimination*, 69 AM. PSYCHOLOGIST 669 (2014).

11. Throughout the Article, we assume that jurors, judges, prosecutors, defense lawyers, and other actors are white and that the suspects or defendants are black. We made this choice for two reasons. First, most jurors, judges, prosecutors, and defense lawyers are white. Similarly, black Americans are the most salient overrepresented group in the criminal justice system—both as victims and offenders. Second, the overwhelming bulk of scholarship on racial disparities in criminal law and procedure make the same assumptions. To be clear, white favoritism is not limited to white actors. A growing body of literature documents "outgroup favoritism," by which, for example, a significant percentage of black Americans also associate white with more positive traits and black with more negative traits. Finally, there is some evidence, though the literature is underdeveloped, that white favoritism extends to preferential treatment of white Americans over Latino- or Asian-Americans, though the degree of the effects and the stability of the findings across contexts are less developed than in the white/black setting.

12. *See infra* notes 57–63 and accompanying text; *see especially* Levinson, *Forgotten Racial Equality*, *supra* note 3, at 354.

13. *See infra* notes 158–168 and accompanying text. *But see* Miles Hewstone, *The 'Ultimate Attribution Error'? A Review of the Literature on Intergroup Causal Attribution*, 20 EUR. J. SOC. PSYCHOL. 311 (1990); Justin D. Levinson, *Mentally Misguided: How State of Mind Inquiries Ignore Psychological Reality and Overlook Cultural Differences*, 49 HOW. L.J. 1 (2005); Thomas F. Pettigrew, *The Ultimate Attribution Error: Extending Allport's Cognitive Analysis of Prejudice*, 5 PERSONALITY & SOC. PSYCHOL. BULL. 461 (1979); Tracie L. Stewart et al., *Consider the Situation: Reducing Automatic Stereotyping Through Situational Attribution Training*, 46 J. EXPERIMENTAL SOC. PSYCHOL. 221 (2010).

14. *See infra* notes 134–149 and accompanying text; *see also* Bobby K. Cheon et al., *Cultural Influences on Neural Basis of Intergroup Empathy*, 57 NEUROIMAGE 642 (2011); Matteo Forgiarini et al., *Racism and the Empathy for Pain on Our Skin*, 2 FRONTIERS IN PSYCOL. 1 (2011); Jennifer N. Gutsell & Michael Inzlicht, *Intergroup Differences in the Sharing of Emotive States: Neural Evidence of an Empathy Gap*, 7 SOC. COGNITIVE & AFFECTIVE NEUROSCIENCE 596, 596 (2012) (finding an in-group bias in empathy); Claus Lamm et al., *How Do We Empathize with Someone Who Is Not Like Us? A Functional Magnetic Resonance Imaging Study*, 22 J. COGNITIVE NEUROSCIENCE 362 (2010); Vani A. Mathur et al., *Neural Basis of Extraordinary Empathy and Altruistic Motivation*, 51 NEUROIMAGE 1468 (2010); Pascal Molenberghs, *The Neuroscience of In-Group Bias*, 37 NEUROSCIENCE AND

white favoritism has serious ramifications for criminal law and procedure because it can operate in a range of powerful ways that can be distinguished from traditional race-focused examples: in the way, for example, white drivers are pulled over less often than unseen drivers, in the way legislators might see white "meth" addicts as suffering from an illness and black "crack" addicts as criminals, and in the way prosecutors and jurors view a crime as more aggravated if the victim is white or see a white juvenile offender to be more capable of redemption.[15] Thus, implicit white favoritism both enriches existing accounts of how implicit racial bias corrupts the criminal justice system and provides explanations for disparities that implicit negative stereotyping explanations miss altogether.

The Article proceeds in three parts. Part I introduces implicit racial bias and explains how scholars have relied on the concept to help explain a range of disparities in the criminal justice system. Despite the importance of the broad model we outline in Part I, we nonetheless conclude that the failure of scholarship to account for implicit favoritism leaves a significant gap in our understanding of why racial disparities persist in the criminal justice system.

Part II begins to address that gap. In this Part, we introduce the social science of implicit favoritism. We begin by exploring how priming positive racial stereotypes—or privileged racial group membership—leads to a boost for racially favored groups. We then investigate how attributions of causation and intentionality are slanted to favor in-group members. We also consider the relative nature of the Implicit Association Test and what it means about implicit favoritism. Finally, we discuss both the fundamentals of enhanced in-group empathy and pain sensitivity and the details of how the human memory favors positively stereotyped groups.

Part III applies the implicit favoritism social science to improve upon the model we developed in Part I. Using the same across-the-system structure we employed in the first Part, we present a series of case studies that illustrate why implicit favoritism is such a powerful tool for understanding disparities in the criminal justice system. We begin with implicit white favoritism and policing, which explores white preference in the decisions to conduct a traffic stop or to shoot an armed suspect. We then consider white favoritism and juries, which includes both the sentencing decision in juvenile life without parole cases and the racial dynamics of self-defense claims. Finally, we explore implicit favoritism

---

BIOBEHAVIORAL REV. 1530 (2013); Stephanie D. Preston & Frans B. M. de Waal, *Empathy: Its Ultimate and Proximate Bases*, 25 BEHAV. & BRAIN SCI. 1 (2002); Xiaojing Xu et al., *Do You Feel My Pain? Racial Group Membership Modulates Empathic Neural Responses*, 29 J. NEUROSCIENCE 8525 (2009).

15. *See infra* Part III.

and legal professionals, which includes legislators, judges, prosecutors, and defense lawyers.

I. A MODEL OF IMPLICIT RACIAL BIAS: BLACK DEROGATION

It is a truism that racial disparities exist in the criminal justice system.[16] These disparities are visible from the initial decision to criminalize conduct through decisions to stop, arrest, prosecute, and sentence citizens. While most commentators agree that these racial disparities are not predominately a consequence of purposeful discrimination, there is comparatively little consensus on why these disparities continue to infect the criminal justice system so thoroughly.[17]

Over the past decade legal scholars have drawn on an extensive body of social science that demonstrates how individual actors in the criminal justice system—and in society generally—possess implicit racial biases that can affect their perceptions, judgments, and behaviors. This Part provides both a thorough overview of the science of implicit racial bias and a comprehensive review of the scholarship that seeks to explain how implicit racial bias can operate at various points of discretion in the criminal justice system. It illustrates that legal scholars have thus far focused on implicit bias as black derogation. Thus, the goal of Part I is to provide necessary background for understanding the importance of grappling with the other half of the implicit racial bias research—implicit bias as white favoritism.

A. An Overview of Implicit Racial Bias

The most efficient way for us to find out your favorite ice cream flavor is simply to ask. Whether it's mint chocolate chip or caramel brownie, there is little reason to worry that your answer will be inaccurate. The direct approach is not the best option for every context, however. Questions about racial preferences and animosities do not lend themselves to direct

16. *See supra* note 1.

17. Some scholars explain the disparities by claiming that blacks are simply more likely to commit crimes as a consequence of a variety of structural and institutional factors—poverty, low levels of education, and high rates of joblessness, for example. Other scholars focus on disparities in existing institutional and structural arrangements that have roots in past formal and intentional racism—the old structures persist, necessarily "reinforc[ing] a racial order where whites are privileged over other groups." Ruth D. Peterson & Lauren J. Krivo, *Race, Residence, and Violent Crime: A Structure of Inequality*, 57 U. KAN. L. REV. 903, 903 (2009); *see also* Gary Ford, *The New Jim Crow: Male and Female, South and North, from Cradle to Grave, Perception and Reality: Racial Disparity and Bias in America's Criminal Justice System*, 11 RUTGERS RACE & L. REV. 324, 329 (2010). More radically, some scholars have claimed that racial discrimination in the criminal justice system is a deliberate and continuous modern mechanism for control of black Americans—a "New Jim Crow."

questioning for two different reasons. First, even if many people continue to hold racist beliefs, Americans have learned that—unlike your favorite ice cream flavor—there is a strong societal commitment to racial colorblindness.[18] Publicly acknowledging a racial preference (or animosity) is socially unacceptable.

The second reason why direct questioning is not effective when asking about racial attitudes is that while many Americans genuinely believe themselves to be colorblind and egalitarian, our minds betray our best intentions.[19] Our minds automatically sort incoming information into categories. This cognitive process is known generally as implicit social cognition.[20] In an attempt to familiarize implicit social cognition for legal audiences, Professor Kang explained its operation with an elegantly simple example: "When we see . . . something with a flat seat, a back, and some legs, we recognize it as a 'chair.' . . . [W]e know what to do with an object that fits into the category 'chair.' Without spending a lot of mental energy, we simply sit."[21] The impact of implicit social cognition is not always so innocuous. Social psychologists have documented the tendency for people to exhibit implicit preferences for groups with higher social status to groups with lower social status.[22] For example, people tend to prefer young people to old people, heterosexual people to homosexual people, abled people to disabled people, and white people to black people.[23]

---

18. *See, e.g.*, Evan P. Apfelbaum et al., *Racial Color Blindness: Emergence, Practice, and Implications*, 21 CURRENT DIRECTIONS IN PSYCHOL. SCI. 205, 205 (2012).

19. *Id*. at 205–06.

20. *See especially* Greenwald & Krieger, *supra* note 3, at 945; Justin D. Levinson, *Introduction: Racial Disparities, Social Science, and the Legal System*, *in* IMPLICIT RACIAL BIAS ACROSS THE LAW 1 (Justin D. Levinson & Robert J. Smith eds., 2012) [hereinafter Levinson, *Racial Disparities*].

21. JERRY KANG, IMPLICIT BIAS: A PRIMER FOR COURTS 1 (2009), *available at* http://wp.jerrykang.net.s110363.gridserver.com/wp-content/uploads/2010/10/kang-Implicit-Bias-Primer-for-courts-09.pdf; *see also* ZIVA KUNDA, SOCIAL COGNITION: MAKING SENSE OF PEOPLE 17–18 (1999); Susan T. Fiske & Steven L. Neuberg, *A Continuum of Impression Formation, from Category-Based to Individuating Processes: Influences of Information and Motivation on Attention and Interpretation*, 23 ADVANCES EXPERIMENTAL SOC. PSYCHOL. 1, 4, 23–24 (1990); Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity*, 97 J. PERSONALITY & SOC. PSYCHOL. 17 (2009); Allen R. McConnell & Jill M. Leibold, *Relations Among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes*, 37 J. EXPERIMENTAL SOC. PSYCHOL. 435 (2001).

22. *See generally* Anthony G. Greenwald & Mahzarin R. Banaji, *Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes*, 102 PSYCHOL. REV. 4 (1995); Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J. PERSONALITY & SOC. PSYCHOL. 1464 (1998); John T. Jost et al., *A Decade of System Justification Theory: Accumulated Evidence of Conscious and Unconscious Bolstering of the Status Quo*, 25 POL. PSYCHOL. 881 (2004).

23. Preference, in this context, refers both to implicit attitudes (e.g., positive or negative) and stereotypes (e.g., intelligent; athletic). *See especially* Mahzarin R. Banaji, *Implicit Attitudes Can Be Measured*, *in* THE NATURE OF REMEMBERING: ESSAYS IN HONOR OF ROBERT G. CROWDER 117 (H.L. Roediger, III et al. eds., 2001); Greenwald et al., *Measuring Individual Differences*, *supra* note 22; Jost et al., *supra* note 22.

The strand of implicit social cognition that addresses the process by which incoming information is sorted into racial categories is labeled implicit racial bias.[24] A rich and overlapping literature documents the tendency of Americans to exhibit implicit racial bias in favor of white Americans over black Americans.[25] Though Asian- and Latino-Americans all show consistent implicit preference for white Americans over black Americans, white Americans register the strongest white-over-black implicit preference.[26]

These preferences—or implicit biases—are typically measured by one of two methods. The first is known as "priming."[27] Priming seeks to assess whether and to what degree exposure to a concept or object (e.g., a black face) automatically activates stereotypes (e.g., "black people are hostile") or shapes stereotype-congruent responses to race-neutral prompts (e.g., rating an ambiguous shove as more aggressive). For example, Professors Laurie Rudman and Matthew Lee used music—either rap or pop—to prime study participants in an effort to determine whether the sound of rap music would activate racial stereotypes and lead participants to render more negative judgments about black people.[28] As predicted, listening to rap music not only activated black stereotypes, but also led participants to judge the behavior of black people as more hostile and less intelligent.[29]

---

24. Levinson et al., *Devaluing Death*, *supra* note 7, at 1; Levinson, *Racial Disparities*, *supra* note 20, at 1.

25. *See, e.g.*, Kang, *Trojan Horses of Race*, *supra* note 3; Levinson, *Forgotten Racial Equality*, *supra* note 3, at 345; Elizabeth A. Phelps et al., *Performance on Indirect Measures of Race Evaluation Predicts Amygdala Activation*, 12 J. COGNITIVE NEUROSCIENCE 729 (2000); Laurie A. Rudman & Richard D. Ashmore, *Discrimination and the Implicit Association Test*, 10 GROUP PROCESSES & INTERGROUP REL. 359 (2007); Denise Sekaquaptewa et al., *Stereotypic Explanatory Bias: Implicit Stereotyping as a Predictor of Discrimination*, 39 J. EXPERIMENTAL SOC. PSYCHOL. 75 (2003).

26. Brian A. Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, 18 EUR. REV. SOC. PSYCHOL. 36, 38 (2007) (noting that roughly one-third of black Americans also exhibit implicit white preference, though one-third exhibit black preference, and the remaining one-third exhibit no statistically significant preference).

27. *See especially* Patricia G. Devine, *Stereotypes and Prejudice: Their Automatic and Controlled Components*, 56 J. PERSONALITY & SOC. PSYCHOL. 5, 7–8 (1989) (showing that the consequence of subconscious activation of negative black racial stereotypes is evaluating ambiguous behavior as aggressive); Jennifer L. Eberhardt et al., *Seeing Black: Race, Crime, and Visual Processing*, 87 J. PERSONALITY & SOC. PSYCHOL. 876, 881, 883, 885–87 (2004) (finding that research subjects primed with crime-related words or photographs were drawn to black faces earlier and for longer periods than white faces); Samuel L. Gaertner & John P. McLaughlin, *Racial Stereotypes: Associations and Ascriptions of Positive and Negative Characteristics*, 46 SOC. PSYCHOL. Q. 23, 23 (1983) (presenting the seminal priming study); Kang, *Trojan Horses*, *supra* note 3, at 1509; Laurie A. Rudman & Matthew R. Lee, *Implicit and Explicit Consequences of Exposure to Violent and Misogynous Rap Music*, 5 GROUP PROCESSES & INTERGROUP REL. 133, 133 (2002).

28. Rudman & Lee, *supra* note 27, at 136–39 (specifying that the average participant listened to the music for thirteen minutes).

29. *Id*.

The second method for measuring implicit racial bias is the best-known one: the Implicit Association Test (IAT).[30] The IAT works by asking participants to pair two categories, for example the racial categories of "white" or "black" with a series of positive (e.g., "motivated") or negative (e.g., "lazy") attributes, typically attitudes or stereotypes. The point of the test is to measure how quickly, in milliseconds, people associate white or black people with positive or negative concepts. Pairing speed matters because when mental connections between a person and an attribute are strong the mind is able to categorize the information more efficiently, which leads to shorter response times. Most white people—over 90%—show implicit white over black attitudinal preferences on the IAT.[31] Black Americans—and especially young black males—are implicitly associated with dangerousness, criminality, and violence.[32]

Using these tools, scholars have begun to explore how implicit racial bias influences criminal justice.[33] In particular, legal scholars have focused on implicit negative stereotyping to explain the importance of implicit racial bias research for understanding disparities in the criminal justice system.[34] Consider what can happen to many Americans when they see a black face. In the same way that we see "a flat seat, a back, and some legs" and think "chair," the associations between black Americans and crime, danger, and violence can become activated when Americans are exposed to

30. Jens B. Asendorpf et al., *Double Dissociation Between Implicit and Explicit Personality Self-Concept: The Case of Shy Behavior*, 83 J. PERSONALITY & SOC. PSYCHOL. 380, 382 (2002); Greenwald et al., *Measuring Individual Differences in Implicit Cognition*, *supra* note 22; Anthony G. Greenwald et al., *Understanding and Using the Implicit Association Test: I. An Improved Scoring Algorithm*, 85 J. PERSONALITY & SOC. PSYCHOL. 197 (2003); McConnell & Leibold, *Relations Among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes*, *supra* note 21. Readers can take the test themselves online at the Project Implicit Website. *See* PROJECT IMPLICIT, http://implicit.harvard.edu/implicit/demo (last visited July 29, 2013). For other types of implicit bias measurement tools, see Irene V. Blair, *The Malleability of Automatic Stereotypes and Prejudice*, 6 PERSONALITY & SOC. PSYCHOL. REV. 242, 260–61 (2002). Richard Banks and colleagues point out that criticism of the Race IAT usually rests on normative disagreements about the nature of bias. *See* R. Richard Banks et al., *Discrimination and Implicit Bias in a Racially Unequal Society*, 94 CALIF. L. REV. 1169, 1186–87 (2006). *But see* Samuel R. Bagenstos, *"Rational Discrimination," Accommodation, and the Politics of (Disability) Civil Rights*, 89 VA. L. REV. 825 (2003).

31. Dasgupta, *Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations*, *supra* note 10, at 147 (cataloging studies and their results); Nosek et al., *Pervasiveness and Correlates of Implicit Attitudes and Stereotypes*, *supra* note 26, at 38.

32. *See* Jennifer L. Eberhardt et al., *Looking Deathworthy: Perceived Stereotypicality of Black Defendants Predicts Capital-Sentencing Outcomes*, 17 PSYCHOL. SCI. 383 (2006); Eberhardt et al., *Seeing Black*, *supra* note 27; Phillip Atiba Goff et al., *Not Yet Human: Implicit Knowledge, Historical Dehumanization, and Contemporary Consequences*, 94 J. PERSONALITY & SOC. PSYCHOL. 292 (2008); Greenwald et al., *Measuring Individual Differences in Implicit Cognition*, *supra* note 22; Rudman & Lee, *supra* note 27.

33. *See generally supra* notes 10–14.

34. *Id.*

a black face.[35] Once activated, these implicit associations can color the real-world behavior of judges and jurors, prosecutors and police, commutation boards, and defense counsel as they make countless decisions across the spectrum of discretionary points in the criminal justice system.

In the remainder of this section, we detail how legal scholars have imported and extended the experimental evidence of out-group derogation into the criminal justice system. We focus on three clusters of actors within the criminal justice system: police officers; jurors; and legal professionals, including legislators, judges, prosecutors, and public defenders.

B. Black Derogation and Policing

Police officers exercise broad discretion to stop, question, search, detain, and arrest citizens. This section isolates two points of discretion that reflect the seriousness and variety of judgments that police officers are entrusted to make: first, whether to stop a specific person and then to frisk that individual for weapons; and, second, the split-second decision of whether to shoot (or not shoot) a potentially armed suspect.

The Fourth Amendment permits a police officer to conduct a "stop and frisk" when "specific and articulable facts" lead the officer "reasonably to conclude in light of his experience that criminal activity may be afoot" or the individual is "armed and presently dangerous."[36] When aggregated these ostensibly race-neutral decisions reflect significant racial disparities. For instance, even though black New Yorkers are a statistical minority, more than half of the pedestrians who were stopped—and frisked—in the past decade by New York City police officers are black.[37] One explanation for these disparities, offered by Professor Richardson, is that police officers tend to "interpret ambiguous behaviors performed by blacks as suspicious [and criminal] while similar behaviors engaged in by whites would go unnoticed."[38] In other words, "[n]onconscious stereotype activation" of black Americans as hostile, dangerous and prone to criminality skews how a police officer perceives and assesses a particular situation.[39]

The notion that implicit negative stereotyping of black Americans contributes to racial disparities in police stops finds support in several recent studies. One study, which used actual police officers as participants,

35. Eberhardt et al., *Seeing Black*, *supra* note 27; B. Keith Payne, *Prejudice and Perception: The Role of Automatic and Controlled Processes in Misperceiving a Weapon*, 81 J. OF PERSONALITY & SOC. PSYCHOL. 181 (2001).

36. Terry v. Ohio, 392 U.S. 1, 12, 21, 30 (1968).

37. This data is collected by the New York Civil Liberties Union and is available at http://www.nyclu.org/content/stop-and-frisk-data (last accessed Aug. 4, 2013).

38. Richardson, *Arrest Efficiency and the Fourth Amendment*, *supra* note 5, at 2062.

39. *Id.*

demonstrated that when primed with crime-relevant words (e.g., arrest, shoot, chase) officers located a dot on a computer screen more quickly when it appeared on the half of the screen that had displayed a black face as opposed to a white face.[40] In a second study, participants correctly identified a degraded picture of a crime-relevant object (e.g., a gun, a knife) as it came into focus in fewer frames when primed with a black face than when either primed with a white face or when they did not receive a prime.[41] Finally, in a third study, police officers evaluated stereotypically black faces as looking "more criminal" than less stereotypically black or white faces.[42] These studies demonstrate a strong automatic association between the concept of crime and black Americans. The crime–black associations, in turn, can trigger suspicion or facilitate the assessment of a black American as a "criminal" or as someone more likely to possess a weapon. In short, implicit racial bias can color the decision of an officer to notice a particular person, stop him, and frisk him for weapons.

The risk that implicit racial bias infects the judgment of police officers does not end with the decision to stop and frisk a suspect. Consider the most serious decision that a police officer could be asked to make: Does this suspect present an imminent threat that necessities the use of deadly force? Does the suspect have a gun—or is that shiny object a cell phone?[43] It is a life or death decision that the officer has milliseconds to make.[44] Unfortunately, in a series of studies known as the "shooter bias" paradigm,

40. Eberhardt et al., *Seeing Black*, *supra* note 27. After being primed with a crime-relevant word, officers saw a screen that contained two images—one of a black face and the other of a white face. The faces disappeared from the screen and a dot-probe was placed on either the half of the screen that contained the white face or the half that contained the black face. *Id*.

41. Eberhardt et al., *Seeing Black*, *supra* note 27. After receiving a prime (or not), participants viewed a background image of a dilapidated neighborhood. In the foreground participants saw one of two degraded images come into focus over a series of forty-one frames. The first image was crime relevant—a gun or a knife—and the second image was crime irrelevant—a camera or a book. *Id*. The Eberhardt study corroborated and improved upon an earlier study by Professor Payne, which found that study participants could identify a "tool" more quickly after being primed with a white face, but could identify a gun more quickly after being primed with a black face. *See* Payne, *supra* note 35.

42. Eberhardt et al., *Seeing Black*, *supra* note 27. The study divided the police officer participants into two groups. The first group saw a series of white faces, while the second group saw a series of black faces. A random subset of the officers also performed a test that required them to rate each face based on the degree to which the face was stereotypically black (or white). Another subset of the officers performed a test that required them to ascertain whether each face "looked criminal." As predicted, the more stereotypically black the face, the more the officers rated the face as looking criminal.

43. Joshua Correll et al., *The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals*, 83 J. PERSONALITY & SOC. PSYCHOL. 1314 (2002); *see also* Charles M. Judd et al., *Automatic Stereotypes vs. Automatic Prejudice: Sorting Out the Possibilities in the Payne (2001) Weapon Paradigm*, 40 J. EXPERIMENTAL SOC. PSYCHOL. 75 (2004); Kang, *Trojan Horses of Race*, *supra* note 3, at 1525–28; Payne, *supra* note 35, at 287.

44. *See* Correll et al., *The Police Officer's Dilemma*, *supra* note 43, at 1314.

researchers have demonstrated that implicit racial bias can influence whether—and how quickly—a person fires a weapon.[45]

The shooter bias paradigm is a customized video game in which a person appears on the screen holding either a cell phone or a gun.[46] Participants are instructed to "shoot" as rapidly as possible if the person is holding a gun or to hit the safety (i.e. "not shoot") as rapidly as possible if the person is holding a cell phone. Participants make the decision to shoot more often and more rapidly when a black person holds a gun compared to when a white person holds a gun.[47] Similarly, an analysis of participant errors shows that participants are more likely to accidentally shoot an unarmed man when he is black than when he is white.[48]

Researchers have even employed this "shooter bias" study using police officer participants.[49] Although police officers were able to avoid making significantly more race-based shooting errors, they nonetheless pulled the trigger more quickly when viewing armed black men compared to armed white men, and hit the safety more quickly when viewing unarmed white men compared to unarmed black men.[50] Although one would hope that an officer's reaction time when making a judgment as to the dangerousness of a potential suspect would not depend on race, this research demonstrates how implicit negative stereotyping can influence how police officers respond to even the gravest scenarios.

C. Black Derogation and Juries

The Sixth Amendment provides criminal defendants the right to trial by an impartial jury. This section discusses the ways in which implicit racial bias can seep into how the jury makes its decisions. It includes, for example, (1) whether the defendant is afforded the presumption of innocence and if the jurors hold the prosecution to its burden to prove each element of the offense beyond a reasonable doubt;[51] (2) how jurors assess

45. *Id.*

46. *Id.*

47. *Id.*

48. *Id.*

49. Joshua Correll et al., *Across the Thin Blue Line: Police Officers and Racial Bias in the Decision to Shoot*, 92 J. PERSONALITY & SOC. PSYCHOL. 1006 (2007).

50. One IAT paradigm, the Weapons–Race IAT, may help to explain the findings of the shooter bias studies. This IAT involves participants pairing either European-American and African-American faces with weapons (e.g., an ax, a rifle) or harmless objects (e.g., a briefcase, a calculator). Most people who complete a Weapons–Race IAT exhibit stronger associations between black people and weapons and white people and harmless objects than between white people and weapons and black people and harmless objects.

51. In re Winship, 397 U.S. 358, 361 (1970) (explaining that "[t]he requirement that guilt of a criminal charge be established by proof beyond a reasonable doubt dates at least from our early years as a Nation").

the probative value of ambiguous evidence; and (3) how jurors evaluate the credibility of a self-defense claim.

First, implicit racial bias corrodes the foundational notions that every defendant is presumed innocent and that the jury can find the defendant guilty only if the prosecution proves each element of the offense beyond a reasonable doubt.[52] The simplest, and most damning, evidence of the corrosive effect of implicit bias on the presumption of innocence and the beyond a reasonable doubt standard comes from a version of the IAT designed to measure whether participants exhibited stronger associations between black and "guilty" and white and "not guilty" as opposed to black and "not guilty" and white and "guilty."[53] As the researchers predicted, participants exhibited significantly stronger associations between black people and "guilty" than they did white people and "guilty."[54] In another study, researchers demonstrated that mock jurors paid more attention to the faces of black men than to white men immediately after sitting in an actual jury box and listening to a federal judge give videotaped jury instructions on the presumption of innocence.[55] Considered together with the strong association between black and "guilty," the researchers interpreted their results to indicate that jurors struggled to reconcile the presumption of innocence with stereotypes of black Americans as dangerous, hostile, and prone to criminality.[56]

Race also can influence the way that jurors assess the probative value of ambiguous pieces of evidence.[57] One study found that participants who

52. *See* Levinson et al., *Guilt by Implicit Racial Bias*, *supra* note 3; Danielle Young et al., *Innocent Until Primed: Mock Jurors' Racially Biased Response to the Presumption of Innocence*, PLOS ONE (2014), *available at* http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0092365&representation=PDF.

53. Levinson et al., *Guilt by Implicit Racial Bias*, *supra* note 3.

54. One might surmise that judges should be entrusted to decide guilt and innocence in criminal trials. Judges, too, exhibit implicit racial bias. *See infra* notes 71–74 and accompanying text.

55. *See* Young et al., *supra* note 52, at 2–3. Professor Danielle Young and her colleagues showed mock jurors (sitting in a jury box) videotaped jury instructions delivered by a federal judge. Half of the mock jurors received jury instructions that included the presumption of innocence instructions, while the other half received jury instructions without it. Immediately following the jury instructions, the researchers employed a dot-probe task and found that the presumption of innocence instructions actually shifted mock jurors' attention towards faces of black men (relative to faces of white men).

56. *Id*.

57. *See* Justin D. Levinson & Danielle Young, *Different Shades of Bias: Skin Tone, Implicit Racial Bias, and Judgments of Ambiguous Evidence*, 112 W. VA. L. REV. 307 (2010). Mock juror participants read a short description of a fictional mini-mart robbery. They then viewed a series of photographs from the crime scene. One of the photographs—the prime—displayed a masked assailant pointing a gun across the counter. The assailant's skin color could be determined from the skin tone of his forearm. Half of the participants saw a darker skinned suspect while the other half saw a lighter skinned suspect. Participants then viewed pieces of evidence discovered during the investigation of the crime and were asked whether and to what degree each piece of evidence was probative of guilt. Participants primed with the dark-skinned suspect believed ambiguous evidence to be more probative of guilt than participants who assessed the same evidence after being primed with a light-skinned suspect. *Id.*

read a short description of a fictional mini-mart robbery interpreted ambiguous pieces of evidence as more probative of guilt after viewing a surveillance photo of a dark-skinned perpetrator than they did after viewing a surveillance photo of a light-skinned perpetrator.[58] The participants primed with the surveillance photo of a dark-skinned suspect also evaluated the overall case against the suspect to be more probative of guilt—both on a 0 to 100 scale and under a traditional guilty/not guilty measure.[59] These results are particularly troubling in light of contemporary knowledge on wrongful convictions because the study suggests race can alter guilt or innocence determinations, at least on the margins. Implicit racial bias also can contribute to wrongful convictions by influencing how people remember—and misremember—people and events. For example, one memory test study, which used actual police officers as participants, found that when primed with crime-related words the officers were more likely to falsely identify the wrong black face as the suspect when the incorrect face exhibited more stereotypically black facial features than the target face.[60]

Most crimes are not "whodunits," however, but richly textured human interactions that turn upon more nuanced *how* and *why* questions. In these cases, the responsibility of juries is to identify gradations of culpability. Consider a second-degree murder case where the accused claims that he acted in self-defense—a claim that often turns on whether the person who exercised self-help had a "reasonable" basis to believe the he was in immediate danger of imminent bodily harm.[61] *Who threw the first punch? What words were exchanged*? *Was either party trying to leave the altercation?* Empirical research demonstrates that jurors remember—and misremember—these types of crime details in racially biased ways, too.[62] Indeed, people who read a fictional story about a confrontation between two men remember aggressive details more accurately when the aggressor has a stereotypically black name—Tyronne.[63] Participants tended also to misremember stereotype-congruent "facts" not included in the story, such as incorrectly remembering Tyronne as the aggressor even when the white man, William, perpetrated the aggression.[64] In addition to these studies

58. *Id*. at 310–11.

59. *Id.* at 337.

60. Eberhardt et al., *Seeing Black*, *supra* note 27, at 888–89.

61. Cynthia Lee, *Making Race Salient: Trayvon Martin and Implicit Bias in a Not Yet Post-Racial Society*, 91 N.C. L. REV. 1555, 1562–63 (2013).

62. Levinson, *Forgotten Racial Equality*, *supra* note 3. Levinson provided participants with a fictional story about a confrontation between two men. Every participant read the same story, except that half of the participants read a story about a white perpetrator named William (who was identified as Caucasian) while the other half read about a black perpetrator named Tyronne (who was identified as African-American). *Id.*

63. *Id*. at 398–400.

64. *Id*. at 400.

demonstrating how racially skewed memory performance can influence self-defense claims, Professor Lee has drawn on the research documenting the associations between blacks and weapons, as well as the shooter bias studies that demonstrate that civilian participants mistakenly shoot unarmed black Americans more often than white Americans, to argue that jurors are more likely to find that the defendant acted reasonably if the deceased was black than if the deceased was white.[65]

D. Black Derogation and Legal Professionals

This Part explores how police officers and jurors exhibit implicit racial bias at multiple stages of the criminal justice system. What about legal professionals—the legislators, judges, prosecutors, and public defenders that define crimes and preside over, prosecute, and defend criminal cases? Here, too, scholars assert the corrosive role of implicit racial bias.

1. Legislators

Implicit bias can enter into the criminal justice system at the very earliest point: when legislators set forth the behaviors that will constitute substantive crimes and set the sanctions that follow from commission of those crimes. We focus on one example: the massive punishment disparity between powder and crack cocaine in the late 1980s.[66]

Debate over the appropriate punishment for crack cocaine occurred in the broader context of the decline of urban America and a rise in inner-city violent crime.[67] Indeed, Americans increasingly associated crack cocaine with "crack murders" and other violent crime.[68] Professor Ogletree and colleagues posit that given this historical backdrop, implicit racial bias—as opposed to simply an explicit intent to disproportionately punish black Americans—contributed to the crack–powder disparity.[69] Americans saw crack as a "black" problem. The link between crack cocaine and blackness triggered automatic associations between black Americans and violence, dangerousness, and criminality. In turn, these associations triggered a fear response. Thus, the "fear of black people ratcheted up the perceived dangerousness of crack cocaine use" and "justified the huge disparity

65. Lee, *Making Race Salient*, *supra* note 61, at 1582–85.

66. Deborah Ahrens, *Methademic: Drug Panic in an Age of Ambivalence*, 37 FLA ST. U. L. REV. 841, 855–57 & n.66 (2010) (noting "crack babies"); *id*. at 853–57 & n.60 (on "crack murders").

67. *Id*. at 857.

68. *Id*. at 855–57 & n.66 (noting "crack babies"); *id*. at 853–57 & n.60 (noting "crack murders").

69. Charles Ogletree et al., *Criminal Law: Coloring Punishment: Implicit Social Cognition and Criminal Justice*, *in* IMPLICIT RACIAL BIAS ACROSS THE LAW 45, 51 (Justin D. Levinson & Robert J. Smith eds., 2012) [hereinafter Ogletree et al., *Coloring Punishment*].

between sentences for the use of crack and powdered cocaine."[70] To be clear, the point made by Ogletree and colleagues is not that implicit racial bias is *the* reason why harsh penalties for drug crimes exist, but rather that equating drug crime—especially crack cocaine usage—with black crime activates a host of negative stereotypes about black Americans and those associations create a filter through which legislative decisions are made.

2. Judges

Judges are the archetype of impartial actors in the criminal justice system. Yet, one recent study casts doubt on the idea that judges are impervious to racial bias.[71] Professor Rachlinski and his colleagues found that white judges from three distinct regions of the United States displayed a strong implicit white-over-black preference on the traditional Race IAT.[72] These researchers then demonstrated that when subliminally primed with words representing stereotypically black concepts, those judges who had demonstrated white-over-black preference on the IAT tended to sentence racially ambiguous defendants more harshly than when they were primed with random words.[73] These findings, though relatively moderate in magnitude, held across two different fictional crime narratives: one involving a juvenile defendant in a violent armed robbery case and the other a juvenile defendant in a non-violent shoplifting case.[74] Though the study authors did not endeavor to interpret their findings, the standard derogation narrative imports easily: black Americans are stereotyped as dangerous, hostile, and prone to criminality; thus, whether rooted in deterrence theory or retribution, black offenders warrant enhanced punishment.

3. Prosecutors

Empirical studies of the federal criminal justice system suggest that prosecutorial discretion—not judicial discretion—is the major driver of

70. *Id.*

71. *See* Jeffrey J. Rachlinski et al., *Does Unconscious Racial Bias Affect Trial Judges?*, 84 NOTRE DAME L. REV. 1195 (2009).

72. *Id.* at 1210.

73. *Id.* at 1214–17.

74. *Id*. Interestingly, in a variation of the study that explicitly identified the race of the defendant, white judges corrected for their implicit bias. After receiving explicit race variation of the study, the judges self-reported suspecting that the study authors intended to test race effects. Hence, the authors concluded that with sufficient motivation judges could overcome their implicit biases.

racial disparities.[75] Most racial disparities in the federal system originate from initial charging decisions, and, specifically, in the decision of whether to charge an offense that carries a mandatory minimum sentence.[76] Imagine a prosecutor deciding whether to charge a juvenile in the adult system as opposed to the juvenile justice system. Prosecutorial guidelines for deciding whether to charge a juvenile as an adult stress the importance of several factors, including (1) "[t]he sophistication and maturity of the juvenile as determined by consideration of his home, environmental situation, emotional attitude and pattern of living";[77] (2) "[t]he seriousness of the alleged offense";[78] (3) "[w]hether the alleged offense was committed in an aggressive, violent, premeditated or willful manner";[79] and (4) "[t]he prospects for adequate protection of the public and the likelihood of reasonable rehabilitation of the juvenile."[80] Are these factors race-neutral or instead influenced by negative stereotypes of black Americans?

Imagine a case where the prosecutor must decide whether to charge as an adult a sixteen-year-old black male who steals candy from a convenience store and punches the clerk in the face before he exits. How culpable is the suspect? Did he punch the clerk in a calculated attempt to avoid being caught, or because he became violent so readily, or simply because he panicked?[81] Professors Smith and Levinson have argued that the race of the juvenile—whether conveyed by a mug shot, a notation in the police file, or even by a stereotypically black name[82]—can trigger stereotypes of black Americans as hostile and aggressive.[83] In turn, "[t]he activation of these negative constructs can translate into a sense that the crime . . . is more aggressive or violent than . . . if the prosecutor [had] assessed the facts of the case in a truly race-neutral manner."[84]

Two empirical studies directly support the hypothesis that implicit racial bias can influence the decision of a prosecutor to charge a juvenile offender as an adult instead of referring him to juvenile court. The first

75. Sonja B. Starr & M. Marit Rehavi, *Racial Disparity in Federal Criminal Charging and Its Sentencing Consequences* (Program in Law & Econ. Working Paper Series, Paper No. 12-002, 2012), *available at* http://www.fjc.gov/public/pdf.nsf/lookup/NSPI201213.pdf/$file/NSPI201213.pdf.

76. *Id*. at 3 ("[Most] sentence gaps can be explained by prosecutors' initial charging decisions—particularly the choice to bring mandatory minimum charges. The mandatory minimum indicator can explain nearly seven percentage points of the race gap at the mean and has a significant and sizeable explanatory effect at every decile, ranging from five to nine percentage points . . . .").

77. Kent v. United States, 383 U.S. 541, 566–67 (1966) (listing factors to be considered in deciding whether to transfer a case to adult court).

78. *Id.* at 566.

79. *Id.* at 567.

80. *Id.*

81. Smith & Levinson, *supra* note 6, at 812.

82. *Id*.

83. *Id*.

84. *Id*.

study found that when primed with words that represent stereotypically black concepts (e.g., "Harlem," "dreadlocks," "homeboy") both police officers and probation officers evaluated juvenile suspects as *more* culpable and deserving of punishment and *less* immature on account of their adolescence.[85] The second study found that participants expressed increased support for a life without parole sentence for a fourteen-year-old defendant who murdered an elderly woman when the researchers described the juvenile as a black male than when they described him as a white male.[86] Moreover, these same participants indicated a stronger belief that juveniles and adults are similarly blameworthy for their conduct when the researchers presented the fourteen-year-old defendant as a black male.[87] Taken together, then, these studies support the notion that implicit racial bias can influence how prosecutors exercise their discretion to charge a juvenile offender in adult court.

4. Public Defenders

What if defense lawyers, too, are susceptible to implicit racial bias? Professors Eisenberg and Johnson used a pencil and paper version of the Race IAT to demonstrate that lawyers who represent defendants facing the death penalty exhibit the same implicit white-over-black preference that exists in the general participant population.[88] Professors Richardson and Goff recently explored how this finding translates into the work of public

85. Sandra Graham & Brian S. Lowery, *Priming Unconscious Racial Stereotypes About Adolescent Offenders*, 28 LAW & HUM. BEHAV. 483, 483 (2004). This study borrowed some of its methodology from research conducted by Patricia Devine. *See* Devine, *supra* note 27, at 7, 9, 13. Professors Graham and Lowery conducted an experiment designed to test whether police officers and probation officers would indicate a preference for harsher punishment after being primed with stereotypically black words. Participants—police and probation officers—read two different fictional crime reports. Both reports described a racially ambiguous juvenile suspect. The first report described an accusation by a convenience store manager that a twelve-year-old boy with no prior record took miscellaneous toys valued at $40 from the store. The boy denied stealing the toys. No witnesses had emerged. In the second crime report, which involved an assault by a fifteen-year-old boy on a sixteen-year-old, conflicting witness accounts made it unclear who started the fight and whether the fifteen-year-old suspect acted in self-defense. Graham & Lowery, *supra* note 85.

86. Aneeta Rattan et al., *Race and the Fragility of the Legal Distinction Between Juveniles and Adults*, PLOS ONE (2012), at 2, *available at* http://www.plosone.org/article/fetchObject.action?uri=info:doi/10.1371/journal.pone.0036680&representation=PDF.

87. *Id*. at 2. Dr. Rattan and colleagues presented participants with a case of a fourteen-year-old juvenile defendant who had brutally raped an elderly woman. Half the participants read the narrative with the juvenile described as being black; the other half read the same narrative except the juvenile was described as white. Subsequently, participants were asked two questions: "'[t]o what extent do you support life sentences with no possibility of parole for juveniles when they have been convicted of serious violent crimes (in which no one was killed)?'" and "'[h]ow much do you believe that juveniles who commit crimes such as these should be considered less blameworthy than an adult who committed the same crime?'" *Id.*

88. Theodore Eisenberg & Sheri Lynn Johnson, *Implicit Racial Attitudes of Death Penalty Lawyers*, 53 DEPAUL L. REV. 1539 (2004).

defender offices generally.[89] With crushing caseloads, public defenders often are forced to "triage" cases,[90] which means that despite best intentions, it is nearly impossible for public defenders to provide adequate representation to every client. How does a lawyer choose to spend more time on one case over another when those cases are very similar in terms of facts and charges? After reviewing studies that Goff conducted, which demonstrate that implicit race bias influences how much punishment participants consider to be justified, Richardson and Goff concluded that "defenders may be more accepting of higher sentencing recommendations for black versus white clients and, thus, less likely to negotiate aggressively for lower sentences or to conduct mitigation investigations."[91] Moreover, "[b]ecause of their belief that a tougher sentence is appropriate or likely to be imposed, [public defenders] may be less likely to fight for their client's release on bail and spend time, effort, and scarce resources negotiating a better plea deal."[92]

***

Taken together, these studies on implicit racial bias illustrate that race continues to matter across the entire criminal justice spectrum, though in ways equally less visible and more pervasive than previously imagined. Yet, our knowledge of the role of implicit bias in criminal law and procedure is constrained. Although the social science literature demonstrates that implicit racial biases can be attributed to derogation (e.g., "black people are hostile"), or favoritism (e.g., "white people are competent"), or both, legal scholars tend to treat implicit racial bias as if it dealt exclusively with the racial derogation that flows from implicit negative stereotyping of black Americans. In Part II, then, we begin to

89. Richardson & Goff, *supra* note 8, at 2626.

90. *Id*. at 2636.

91. *Id.* at 2640. Goff himself helped to pioneer the research on implicit dehumanization. In an emotionally powerful series of studies, Goff and colleagues found that implicit racial biases can strike at the core of what it means to be fully human. In one study, participants recognized degraded pictures of apes in fewer frames when primed with a black face than when primed with a white face. In another study, participants primed with an ape were faster to detect a dot-probe when positioned where a black face had appeared on a screen than when the dot-probe was positioned where a white face had appeared on the screen. After establishing this bi-directional relationship between "black" and "ape," Goff and colleagues had participants in a third study watch a video of a police chase that ended with police officers beating the black suspect after he had exited his vehicle. Use of an ape prime (but not a feline prime or no prime at all) significantly increased the degree to which participants believed that the officers were justified in their harsh treatment of the black suspect. Richardson and Goff's article translates these findings into the public defender context. *See also* Smith & Cohen, *supra* note 7 (suggesting that the black–ape studies suggest that implicit dehumanization could influence the relationship between capital defenders and white clients differently than capital defenders and their black clients).

92. Richardson & Goff, *supra* note 8, at 2641.

build a more accurate model of implicit bias-driven inequality by introducing the concept of implicit white favoritism—a sort of automatic boost that one receives on account of his whiteness. Our favoritism account does not supplant the derogation account. To the contrary, it illustrates why criminal law and procedure scholars concerned about racial disparities in the criminal justice system must grapple with both derogation and favoritism.

II. THE SOCIAL SCIENCE OF IMPLICIT FAVORITISM

This Part introduces the equally established other half of implicit racial bias: implicit white favoritism.[93] Implicit white favoritism refers to the positive effects of implicit bias on members of privileged groups. Understanding the mechanics of implicit favoritism is important for anyone who seeks to reduce disparities in the criminal justice system. Removing black derogation is not the same as being race-neutral. Even if all implicit negative stereotyping miraculously ended tomorrow, unjustified racial disparities would persist for at least as long as legislators, police officers, jurors, and legal professionals continue to engage in white favoritism. In Part III we explore concrete examples of how implicit favoritism perpetuates racial disparities in criminal justice. First, though, we explore the science of white favoritism.

This Part begins by describing how priming positive racial stereotypes leads to positive gains for white Americans. Next, we investigate how attributions of causation and intentionality are slanted to favor in-group members, who, in the criminal justice system, are white Americans. We then consider the relative nature of the IAT and what it means for implicit favoritism. Finally, we discuss the fundamentals of enhanced in-group empathy and detail how memory errors may favor white Americans. The purpose of this Part is not only to illustrate implicit racial bias as implicit favoritism, but also to demonstrate that favoritism itself is an umbrella term that captures a variety of ways in which Americans tend to implicitly advantage whites.

A. Priming: A Powerful In-Group Boost

The research on priming discussed in Part I demonstrated that when stereotypes related to blacks are activated, these stereotypes could have severe consequences for their targets. But this implicit derogation effect is only half of the story. The other half of the story is that positive stereotypes

93. Marilynn B. Brewer, *In-Group Bias in the Minimal Intergroup Situation: A Cognitive–Motivational Analysis*, 86 PSYCHOL. BULL. 307 (1979); Greenwald & Pettigrew, *supra* note 10.

of whites lead to broad and significant benefits, which include an enhancement in the way that others view the stereotyped targets and elevations in their own enhanced performance.

1. Stereotype Boost

One of the most famous social psychological priming methodologies occurred not in the criminal stereotype realm, but in the context of academic test performance. Claude Steele and Joshua Aronson primed Caucasian and African-American college students by asking them to identify their race just before they took a test.[94] The researchers found that such a simple priming task had profound negative effects on African-American test performance: African-American participants took longer to answer questions and achieved lower overall scores than Caucasian participants, but only when they were primed.[95] Thus, Steele and Aronson found that priming a participant's racial identity likely implicated a complex relationship between African-American identity and the activation of negative stereotypes relating to ability.[96] They called this phenomenon "stereotype threat."[97] As scholars, we have told this stereotype threat story from the perspective of how negative stereotypes of black Americans can lead to serious harms.[98] Though that story is accurate and deeply troubling, it leaves out the effects of favoritism: First, the activation of positive stereotypes can surprisingly lead to a boost for the target of the stereotype. Second, activating negative stereotypes of out-groups can "lift" the performance of in-group members.

Researchers have found that indirectly activating positive stereotypes about favored group members enhances positive stereotype-consistent performance by members of those groups.[99] Margaret Shih and her colleagues, for example, primed the ethnic identity of Asian-American

94. Claude M. Steele & Joshua Aronson, *Stereotype Threat and the Intellectual Test Performance of African Americans*, 69 J. PERSONALITY & SOC. PSYCHOL. 797 (1995).

95. *Id.*

96. *Id.* at 797–98.

97. *Id*.

98. Justin D. Levinson et al., *Implicit Racial Bias: A Social Science Overview*, *in* IMPLICIT RACIAL BIAS ACROSS THE LAW 9 (Justin D. Levinson & Robert J. Smith eds., 2012).

99. Margaret Shih et al., *Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance*, 10 PSYCHOL. SCI. 80 (1999). It is notable that Shih and her colleagues used indirect priming methods to activate the positive stereotype. Some research has indicated that direct activation of stereotypes in some circumstances (such as when there is pressure or observers are watching) may not trigger the same effects. Margaret Shih et al., *Stereotype Performance Boosts: The Impact of Self-Relevance and the Manner of Stereotype Activation*, 83 J. PERSONALITY & SOC. PSYCHOL. 638 (2002); *see also* Anne C. Krendl et al., *The Effects of Stereotypes and Observer Pressure on Athletic Performance*, 34 J. SPORTS & EXERCISE PSYCHOL. 3 (2012).

female study participants[100] and hypothesized that doing so would activate positive stereotypes related to Asians' math performance (compared both to participants' whose female identity had been primed and to a control group). This stereotype boost hypothesis was confirmed.[101] The priming of Asian identity activated participants' positive stereotypes, and in turn, led to improved math scores on standardized test-type measures.[102] The study, then, demonstrated that the activation of ethnicity-related positive stereotypes can lead to enhanced performance.[103] In the criminal justice system, such positive stereotypes rest predominantly with whites and other favored groups.

Stereotype boost has also been shown to function beyond the academic domain, such as in providing white athletes with a performance bump. For example, white participants in a study who were told that white basketball players were the best free throw shooters in the National Basketball Association (NBA) actually made more free throws than those in a control group as well as those in a group told that black NBA players were the best free throw shooters.[104] Despite their close connection, both in methodology and in the association between stereotype and stereotype-consistent performance, researchers believe that stereotype boost is an independent cognitive phenomena from stereotype threat.[105] That is, the boost experienced by members of positively stereotyped groups is not simply the opposite effect of the negative effects experienced by members of negatively stereotyped groups.[106] Instead, the boosting function serves

100. The participants' Asian identity was primed indirectly. Participants in the ethnicity priming condition were asked questions about what languages they spoke at home, whether they were members of any student organizations, and how many generations of their family had lived in America. Shih et al., *Stereotype Susceptibility*, *supra* note 99, at 80–81.

101. *Id.* at 80.

102. *Id*.

103. *Id.* In a follow-up study, Shih and her colleagues primed their all-female study participants by asking them indirect questions about their gender identity. Margaret J. Shih et al., *Stereotype Boost: Positive Outcomes from the Activation of Positive Stereotypes*, *in* STEREOTYPE THREAT: THEORY, PROCESS, AND APPLICATION 41 (Michael Inzlicht & Toni Schmader eds., 2011) [hereinafter Shih et al., *Positive Outcomes*]. The researchers found that these gender-related primes actually activated the positive stereotypes about women and verbal ability, and led to increased performance on verbal test scores compared to a control group (as well as compared to participants whose Asian identity had been primed). One of the more interesting parts of Shih and colleagues' various studies is that they discovered that, even in groups of people with multiple identities (such as Asian-American females), stereotype boost can be found by priming the relevant identity associated with the positive stereotype. *Id.*

104. Krendl et al., *The Effects of Stereotypes and Observer Pressure on Athletic Performance*, *supra* note 99. In each of these stereotype boost studies, whether in the test-taking realm or beyond, researchers have determined that, in order to trigger the boost, the activation should be implicit. That is, the positive stereotype must be activated indirectly (and thus without increased pressure on the target of the stereotype, such as telling the free throw shooter that they will be videotaped for purposes of posting the video on YouTube in order for the boost to occur. *Id.*

105. Shih et al., *Positive Outcomes*, *supra* note 103.

106. *Id.*

important, ego-related purposes that allow the actor and other in-group members to enjoy the fruits of one's well-positioned group membership. This distinction is analogous to the broader implicit bias discussion, such that we need not consider out-group derogation and in-group favoritism as analogous concepts. Rather, each phenomenon may serve important and sometimes independent cognitive and psychological purposes.

2. Stereotype Lift

Stereotype boost can be distinguished from another phenomenon that leads to in-group gains, known as "stereotype lift." Unlike stereotype boost, which results directly from the activation of positive stereotypes of group members, stereotype lift describes the bump that non-stereotyped members get simply by the activation of negative stereotypes about out-group members.[107] This interesting effect has been shown to be particularly consistent in the academic testing world, where white students (and men, in particular) perform better when a negative out-group stereotype has been activated compared to no stereotype activation at all.[108] As social psychologists Greg Walton and Cohen describe, "[C]omparing themselves with a socially devalued group, people may experience an elevation in their self-efficacy or sense of personal worth, which may, in turn, improve performance."[109] Interestingly, stereotype lift is simple to activate in the testing setting. Because negative stereotypes regarding the academic ability of minorities are so strong, simply describing a test as diagnostic can elicit knowledge of these out-group derogative stereotypes and trigger a corresponding lift in in-group members.[110] Stereotype lift research thus demonstrates that not only do members of positively stereotyped groups get a boost from the indirect activation of those stereotypes, as we have described, but also separately from the activation of negative stereotypes of out-group members. Both theories help provide context for how implicit in-group favoritism works both in and out of the criminal justice system.

Taken together, the research on stereotype boost and stereotype lift shows how implicit favoritism and the existence of positive stereotypes towards members of privileged groups can work on a non-conscious level and lead to various forms of automatic self-enhancement by members of

107. *Id.*

108. Gregory M. Walton & Geoffrey L. Cohen, *Stereotype Lift*, 39 J. EXPERIMENTAL SOC. PSYCHOL. 456 (2003).

109. *Id.* at 456 (internal citation omitted) (citing ALBERT BANDURA, SOCIAL FOUNDATIONS OF THOUGHT AND ACTION: A SOCIAL COGNITIVE THEORY (1986)).

110. *Id.*

those groups.[111] These benefits to the self, however, are matched in the priming context by research that delivers positive priming benefits beyond the self, to other members of privileged groups.

3. An "Us" Versus "Them" Implicit Mentality

At the core of research on implicit in-group favoritism is the principle that people automatically associate the in-group, or "us," with positive characteristics, and the out-group, or "them," with negative characteristics. In a creative study that was one of the first to use implicit measures in the context of in-group favoritism literature, Charles Perdue and his colleagues set out to test just how automatically people associate the in-group with positive traits.[112] In their study, the researchers presented participants with nonsense syllables (sounds without real meaning, such as "xeh," "yof," or "laj") and repeatedly grouped these meaningless syllables together with either in-group or out-group descriptor words, such as "us" and "them," or "we" and "they."[113] The researchers hypothesized that when the meaningless syllables were repeatedly associated with either in-group or out-group descriptor words, these syllables would not only become cognitively associated with in-group or out-group, but would actually become to be viewed more positively or negatively by participants.[114] This hypothesis was indeed confirmed. Participants rated meaningless syllables more positively when they had been grouped with in-group descriptors and more negatively with out-group descriptors.[115]

Building on their initial study, the researchers next set out to test whether in-group and out-group descriptor words, subliminally primed, could actually influence the speed at which positive and negative concepts were recognized.[116] In that study, participants were subliminally primed by flashing in-group and out-group descriptor words (on a computer screen) at speeds at which they were unrecognizable but would nonetheless register subliminally.[117] The words were immediately followed by either positively

111. In addition to stereotype boost and stereotype lift, social psychologists have identified other related mechanisms that help maintain the success and self-esteem of in-group members in particular situations. Nilanjana Dasgupta, for example, has argued that "ingroup members (experts and peers in high-achievement settings) function as 'social vaccines' who increase social belonging and inoculate fellow group members' self-concept against stereotypes." Dasgupta, *Ingroup Experts and Peers*, *supra* note 10, at 231 (italics omitted).

112. Charles W. Perdue et al., *Us and Them: Social Categorization and the Process of Intergroup Bias*, 59 J. PERSONALITY & SOC. PSYCHOL. 475 (1990).

113. *Id*.

114. *Id.* at 476.

115. The results also indicated that the in-group positive associations occurred in the in-group category compared to the control group. *See id*. at 477.

116. *Id.* at 478.

117. *Id.* at 478–79.

or negatively valenced traits, and the researchers measured the speed at which people were able to categorize these traits as being either positive or negative.[118] The researchers found that simply priming in-group words allowed participants to recognize and process positive words faster, whereas priming out-group words allowed participants to recognize and process negative words faster.[119] This finding was one of the initial confirmations that even non-consciously activating in-group constructs could facilitate the processing of positive concepts. One limitation of that study, particularly for the purposes of declaring it to be a perfect indicator of in-group favoritism, is that the researchers could not disambiguate whether in-group words facilitated positive processing, whether out-group words facilitated out-group derogation, or whether it was a combination of the two. Or put more simply, their finding, although powerful, was relative such that it was unclear how much of it was due to in-group favoritism alone.

Fortunately, another study by the same researchers addressed this challenge.[120] Consistent with the principles of implicit in-group favoritism, the study confirmed that the strongest priming effect was consistent not with out-group derogation, but instead with in-group favoritism. In that study, the researchers essentially repeated the core elements of the second study, but added a control group that was subliminally primed with a meaningless letter string, like "xxx."[121] Adding this control group allowed the researchers to determine whether or not the effects they found in the prior study were primarily driven by in-group favoritism, out-group derogation, or a mix of the two. Interestingly, results of the study showed that participants primed with in-group words were significantly faster to respond to positive traits than both control and out-group members. In addition, they found that both the control group and the out-group tended to respond more slowly to positive traits.[122] Thus, the researchers concluded that the in-group favoritism they found was a separate and strong phenomenon.[123] The study, thus, set the stage for additional studies focusing on the automatic categorization effects of in-group membership,

118. *Id.*

119. *Id.* at 479–80.

120. *Id.* at 480.

121. *Id.*

122. *Id.* at 481–82.

123. *Id.* For similar study methodologies, see John A. Bargh & Paula Pietromonaco, *Automatic Information Processing and Social Perception: The Influence of Trait Information Presented Outside of Conscious Awareness on Impression Formation*, 43 J. PERSONALITY & SOC. PSYCHOL. 437 (1982). *See also* Brewer, *supra* note 93.

including the cognitive effects of priming constructs associated with the racial category of white.[124]

4. Priming "White": Activating Positive Stereotypes and Inhibiting the Recognition of Weapons

Building on research that demonstrated how group membership facilitates the implicit transmission of in-group favoritism, researchers began priming participants directly with implicit racial constructs and measuring the effects of these primes. One such study was conducted by Bernd Wittenbrink and his colleagues, who subliminally primed white American participants with the words "white," "black," a control group non-word prime ("XXXXX"), or a filler word prime (e.g., "lemon").[125] Immediately following the prime, researchers presented participants with a series of word recognition tasks known as "lexical-decision task[s]," in which participants were asked to recognize words (or recognize them as non-words) as quickly as possible.[126] These words had been designed to be either stereotypical (positive or negative) of whites and blacks.[127] The speed of the participants' recognition of the word or non-word was then compared based upon the type of prime (white/black/control) and the positive or negative valence of the stereotypic words presented. Results of the study showed that racially priming participants automatically made stereotype-consistent information cognitively available.[128] Consistent with an out-group derogation model, those who were primed with black recognized negative stereotype words more quickly than those primed with white (or with a control).[129] And, consistent with implicit in-group

124. A follow-up study conducted by Pratto and Shih replicated Perdue and colleagues' paradigm, but also investigated the research question of whether the "us" versus "them" implicit mentality might vary based upon the participants' individual preferences relating to social dominance orientation (using a "social dominance orientation" measure that is similar to a self-report bias scale). *See* Felicia Pratto & Margaret Shih, *Social Dominance Orientation and Group Context in Implicit Group Prejudice*, 11 PSYCHOL. SCI. 515, 515 (2000). The experimenters noted that under normal conditions, there was no difference in terms of in-group bias shown based upon social dominance differences among participants. However, when the experimenters created a universal threat condition, under which the likelihood of an in-group membership being threatened was enhanced, the researchers found that both in-group favoritism and out-group derogation were heightened among participants with a high social dominance orientation. Thus, unlike Perdue and colleagues' study, in which the researchers found that in-group favoritism was the driving force behind their results, Pratto and Shih found that threat conditions independently triggered both strong in-group favoritism and out-group derogation. *Id.* For an explanation of Social Dominance Orientation, see Levinson, *Forgotten Racial Equality*, *supra* note 3, at 360–61.

125. Bernd Wittenbrink et al., *Evidence for Racial Prejudice at the Implicit Level and Its Relationship with Questionnaire Measures*, 72 J. PERSONALITY & SOC. PSYCHOL. 262 (1997).

126. *Id.* at 265.

127. *Id.* at 264.

128. *Id*. at 268.

129. *Id.*

favoritism, those who were primed with white recognized positive stereotype words faster than those primed with black (or with a control).[130] Thus, with respect to the implicit favoritism element of the study, priming white independently triggered a cognitive progress separate from out-group derogation that worked to activate positive white racial stereotypes.

Following Wittenbrink and colleagues' study, Eberhardt and colleagues examined whether subliminal racial priming of white or black faces led to faster or slower identification of weapons.[131] This study is often cited for its compelling out-group derogation effects (as we described in Part I), which indeed raise a host of justice issues. But, as we will explain, the study showed more than implicit out-group derogation related to black stereotypes of violence. It also demonstrated the favoritism that may occur because white citizens are automatically and cognitively disassociated with violence. Specifically, Eberhardt and her colleagues found that, when primed with white faces, participants were not only slower to identify weapons compared to when participants were primed with black faces, but also that they were slower to identify weapons compared to participants who were not primed at all.[132] Although the participants were entirely unaware of whether or not they had even been primed, simply seeing a white face for mere milliseconds made it significantly harder for them to perceive a weapon than when they saw no face at all.[133] This study adds to the priming literature that has demonstrated the various ways that in-group membership automatically leads to powerful and important differences, both in the ways that in-group membership performance is enhanced, the ways in which in-group members' activities may be perceived differently by third parties, and the ways in which priming white activates positive racial stereotypes.

B. More Favoritism: Empathy, the IAT, and Other Implicit Measures

Beyond the various automatic in-group advantages shown by priming methodologies, a variety of other methods have been used to demonstrate the ways in which implicit in-group favoritism operates, all of which may help to explain disparities in the criminal justice system. For example, research in the fields of empathy, attribution, and implicit attitudes and stereotypes all contribute to the proof of powerful, independent, automatic in-group favoritism.

130. *Id*.
131. *See* Eberhardt et al., *Seeing Black*, *supra* note 27.
132. *Id.* at 880.
133. *Id.*

1. Empathy with In-Group Members

Being a member of a socially dominant or positively stereotyped in-group does not only have advantages when it comes to performance boosts, the facilitation of positive associations, and the inhibition of negative associations. It also has equally dynamic effects in terms of the way people are reacted to by others, including powerful interpersonal connections of empathy. In the criminal justice context, it is hard to isolate an instance in which empathy with a suspect, defendant, or victim is unimportant.

Research from social, cognitive, and neuropsychology has found that empathy is experienced more for in-group members than out-group members. Empathy is defined by social scientists as "the ability to understand and vicariously share the feelings and thoughts of other people."[134] These feelings are critical both inside and outside the legal system because, as Matteo Forgiarini and colleagues describe, "they enable human beings to tune their mental states to their social environment as well as to understand others' intentions, actions, and behaviors."[135] In the criminal justice system, police, prosecutors, judges, jurors, and probation officers all must seek to understand such intentions, actions, and behaviors. Thus, if people's empathic abilities are fraught with implicit favoritism for in-group members, one might expect that disparities could result at various stages in the system.[136]

People's empathic reaction to others' pain is one particular component of empathy that is recognized by researchers to be based not on conscious ability or recognition, but instead is largely automatic in nature.[137] Unfortunately, research has repeatedly demonstrated that these automatic reactions do not function equally to all others' pain. In fact, the more similar the "witness" and the person experiencing the pain, the more one can expect an automatic empathic reaction to occur.[138] Several studies have confirmed this finding—and related principles—by examining the

134. Forgiarini et al., *supra* note 14, at 1.

135. *Id.* As Jennifer Gutsell and Michael Inzlicht describe,

> [The] vicarious activation of similar neural networks then elicits the associated autonomic and somatic responses, thereby allowing the observer to share the target's motivational and emotional states. Empathy, thus, comes to us naturally—we "catch" the other's emotions—and this vicarious experience helps us to gain an intuitive understanding of them.

Gutsell & Inzlicht, *supra* note 14, at 596 (finding an in-group bias in empathy) (citation omitted).

136. After all, researchers explain that "lack of empathy for the pain of other human beings may lead to violence, abuse, and deterioration of interpersonal and intergroup relationships." Forgiarini et al., *supra* note 14, at 1.

137. *Id.*

138. *Id.* (citing Dennis Krebs, *Empathy and Altruism*, 32 J. PERSONALITY & SOC. PSYCHOL. 1134 (1975); Stephanie D. Preston & Frans B. M. de Waal, *Empathy: Its Ultimate and Proximate Bases*, 25 BEHAV. BRAIN SCI. 1 (2001); Lamm et al., *supra* note 14.

neurological processes associated with empathy.[139] In a study of affective empathy by Xu and colleagues, the researchers used fMRI (functional Magnetic Resonance Imaging) to measure the elements of what scientists call the "pain matrix": regions in the brain that are activated when a person sees another in pain.[140] They tested whether Chinese and Caucasian participants exhibited different brain region reactions when watching in-group and out-group members in pain (receiving a needle injection into the cheek compared to rubbing a Q-tip on the same cheek).[141] The results of the fMRI scans showed that, relative to the Q-tip touch, needle penetrations led to the activation pain matrix brain regions (the Anterior Cingular Cortext, in particular) for in-group members but not out-group members.[142] Phrased another way, in-group members' pain triggered empathic brain responses in fellow in-group members but not in out-group members.

A 2011 study not only replicated the finding that people physiologically react more to in-group members' pain than out-group members' pain, but also connected this finding to implicit racial bias.[143] In that study, Matteo Forgiarini and colleagues tested the empathy-related skin conductance responses of white (Italian) participants who viewed video clips of white, Asian, or African people in pain. In each video clip, participants saw a person with a neutral facial expression.[144] The camera then shifted perspective and zoomed in on the person's hand, which was either touched by an eraser or a needle.[145] The researchers found, first, that the participants exhibited greater skin conductance responses when watching white people in pain than for African people in pain.[146] The researchers then tested whether race-based empathic differences could be predicted by implicit racial attitudes towards whites and blacks. To that end, they administered a Black–White Attitude IAT to participants and tested the relationship between the IAT scores and the participants' empathic reactions to white and African people in pain.[147] Interestingly, the results of the study demonstrated that implicit racial bias predicted the race

139. *See* Molenberghs, *supra* note 14, at 1533. There are two types of empathy—affective and cognitive. Affective empathy describes people's "ability to resonate with what a person is feeling" while cognitive empathy refers to "the ability to take the perspective of someone else and understand what the person is thinking." *Id.* The study by Xu and colleagues, as we describe, measures affective empathy. Other studies have examined in-group biases in cognitive empathy. *See, e.g.*, Cheon et al., *supra* note 14, at 642; Mathur et al., *supra* note 14, at 1468.

140. These regions are the anterior insula and dorsal anterior cingulated cortex. Molenberghs, *supra* note 14, at 1530; Xu et al., *supra* note 14, at 8525.

141. Xu et al., *supra* note 14, at 8525.

142. *Id.* at 8527.

143. Forgiarini et al., *supra* note 14, at 1.

144. *Id.*

145. *Id.*

146. *Id.*

147. *Id.*

differences in empathic response.[148] The more implicit bias the participants showed, the more likely they were to display race-based empathic response biases in favor of white people.[149] The results of the study confirm not only that race-based differences in empathy for pain exist, but also that such seemingly automatic neurological responses are socially dependent and connected with implicit racial bias.

2. Implicit Memory—Forgetting White Aggressiveness

Social psychology often relies heavily on the role of memory in cognitive and decision-making processes. As such, memory errors have been studied frequently since the 1980s as a source of understanding how bias manifests.[150] In the legal context, the role of memory—for judges, jurors, and eyewitnesses—has been identified as a key factor in potentially understanding the role of racial biases in decision making. Levinson's 2007 empirical study on implicit memory bias, in particular, is cited for the proposition that black defendants are implicitly penalized because jurors are more likely to remember their aggressive actions. In his study, Levinson presented mock jurors with facts from a fight that began as a confrontation at a bar, distracted participants for approximately fifteen minutes, and then tested their recall of the case facts.[151] He found that participants who read about a black defendant (Tyronne) accurately recalled more of the aggressive facts from the case (just over 80%), whereas participants who read about a white defendant (William) recalled fewer such facts (under 70%).[152] Although there was no control group in Levinson's study (the study also tested memory biases related to a Native Hawaiian perpetrator, as we will discuss, but this was not a control group), the methodology suggests that implicit favoritism may have acted to inhibit recall of stereotype-inconsistent information related to a white perpetrator. Such a result would be consistent with Eberhardt's study, in which the activation of white stereotypes inhibited the identification of weapons.[153]

Another finding in Levinson's study of memory errors similarly supports the connection between the activation of positive stereotypes (such as non-violence) and memory errors in the criminal justice system. In addition to measuring accuracy of recall, the study also measured whether

148. *Id.*

149. *Id.*

150. *See* Levinson, *Forgotten Racial Equality*, *supra* note 3, at 348–49.

151. *Id.*

152. *Id.*

153. We acknowledge that it is nonetheless possible that negative stereotypes of aggressive black males drove this finding. Additional research would be required to test it.

participants displayed false memories of favorable mitigating evidence.[154] That is, it tested whether participants incorrectly remembered that the defendant, for example, had been provoked. Interestingly, the results on this measure showed that participants exhibited such false memories of mitigating facts when they read about a Native Hawaiian defendant (more frequently than a black or white defendant).[155] Levinson explained this finding as potentially connecting local stereotypes regarding the provocation of Hawaii's native peoples and its translation to the courtroom: "the false memory result could be explained if the provocation of Hawaiians was a societal stereotype."[156] This interpretation is consistent with social cognition research that indicates that people are more likely to falsely remember expectancy-consistent information.[157]

Although the data on the role of positive stereotypes on courtroom memory errors needs to be further studied in order to separate in-group favoritism from out-group derogation biases, the findings suggest that the activation of positive stereotypes (whites as non-aggressive and Hawaiians as being justified in aggressive response) may lead to memory errors and false memories that benefit members of these positively stereotyped groups.

3. Altered Attributions; or Favoritism in Explaining Why

Like the research on empathy and implicit memory processes, research on the way we understand others' behavior has been linked to in-group favoritism. In this context, the research dates back to 1979, when Thomas Pettigrew first described the "Ultimate Attribution Error."[158] The Ultimate Attribution Error describes the "systematic patterning of intergroup misattributions shaped in part by prejudice."[159] Attributions, of course, are central to criminal law processes, ranging from judgments of causation to evaluations of internal mental state (such as intentionality).[160] Thus,

154. *See* Levinson, *Forgotten Racial Equality*, *supra* note 3, at 394.

155. *Id.* at 402.

156. *Id.*

157. C. Neil Macrae et al., *Creating Memory Illusions: Expectancy-Based Processing and the Generation of False Memories*, 10 MEMORY 63, 65 (2002). This finding has been shown to even be heightened in conditions that detract from the amount of focus employed by the participant, which could be similar to a trial in which massive amounts of information are conveyed, for example.

158. Pettigrew, *supra* note 13.

159. Hewstone, *supra* note 13, at 311. Hewstone, after reviewing the literature on intergroup causal attributions, suggests the "Ultimate Attribution Error" is not the best name for the phenomenon, instead suggesting "'intergroup attributional bias'." *Id.* at 332.

160. *See* Levinson, *supra* note 13, at 1 ("[P]olicymakers develop law related to the human mind without an understanding of the human mind itself.").

automatic in-group favoritism in attributions would be particularly concerning in the criminal justice process.

Since Pettigrew's initial hypothesis was presented, many studies have tested whether perceivers indeed ascribe internal attributions (morality, intentionality) to the successes, and external attributions (extenuating and temporary circumstances), to the shortcomings of in-group members.[161] One large-scale study by Miles Hewstone aggregated a range of studies on intergroup attributions (study participants included American children, American college students, Malaysian citizens, students in Singapore, Arab and Israeli participants, and more) and considered the results together.[162] His examination found that, indeed, there was a pattern whereby people tended to attribute the positive behaviors of in-group members as dispositional and negative behaviors as situational.[163] Similarly, people attributed the positive behaviors of out-group members as situational.[164] However, there were not clear findings that negative behaviors of out-group members were typically categorized as dispositional.[165] With regard to the question (important for purposes of this Article) of whether the attributional differences are based upon more in-group favoritism or out-group derogation, the research tends to suggest that in-group favoritism can be a stronger motivator than out-group derogation, although there is ample evidence of both.[166] Furthermore, Hewstone found that within the category of in-group favoritism, there was stronger evidence of in-group protection (regarding making situational attributions for negative actions) than there was evidence of in-group enhancement (regarding making dispositional attributions for positive actions).[167] Thus, the evidence suggests that one of the strongest potential functions of the Ultimate Attribution Error is to protect the in-group in situations of potential negativity (by shifting attributions from internal to external). It is this type of self and in-group status quo protection that may similarly drive other implicit favoritism phenomena, a possibility that we discuss *infra*.[168]

161. See Hewstone, *supra* note 13, for a description of many of these studies.

162. *Id.*

163. *Id.*

164. *Id.* at 312.

165. *Id.* at 316.

166. *Id.*

167. *Id.* at 322.

168. Although original work on the Ultimate Attribution Error tends not to focus specifically on the question of whether the attributional errors are automatic in nature, more recent research focused on "de-biasing" these errors indicates that such errors indeed are driven by automatic stereotyping processes. *See, e.g.*, Stewart et al., *supra* note 13, at 221.

4. Finding Favoritism in the Implicit Association Test

Much of the legal scholarship on the IAT has tended to focus on the powerful and automatic cognitive links people hold that associate subordinated groups to the negative attitudes and stereotypes about them. Indeed, such findings are powerful in the criminal justice context and beyond, but they should also be considered within the relative context of the test itself. As we have described supra Part II, the standard form of the IAT measures how people categorize group members with positive and negative attitudes or stereotypes. For example, in the classic Race-Based IAT, participants are asked to group together photos of white and black people with words representing either good or bad. A strong implicit association harbored by a test-taker in this case means that it is easier for that test-taker to group together not just black and bad and white and good, but actually that it is easier for that test-taker to group together both black and bad and white and good (combined) compared to black and good and white and bad (also combined). As a result, using the standard IAT methodology alone, it can be difficult to determine whether the results are driven only by a strong association between black and bad, a strong association between white and good, or both. For purposes of understanding how existing scholarship on racial inequality can be complemented with a broader understanding of implicit favoritism, it is important to understand the relative nature of the IAT and what it means.

Consider the interpretational challenge posed by the "Ideal Litigator IAT" developed by Kang and colleagues.[169] In that study, the researchers found that people implicitly associated litigators with white and scientists with Asian.[170] Note that from the results of the study, one cannot accurately determine which association is stronger. As Nosek and Banaji observed more generally, "[t]he structure of the IAT constrains evaluations to be relative comparisons between two opposing categories."[171] In fact, it could be possible that there is no particular association between litigator and white compared to litigator and Asian, but instead that people have particularly strong associations of Asian and scientist compared to white and scientist. Similarly, the Guilty/Not-Guilty IAT designed by Levinson and colleagues found that people implicitly associate black with guilty and white with not guilty, compared to black with not guilty and white with guilty.[172] These results are powerful either way, but one cannot determine

169. Jerry Kang et al., *Are Ideal Litigators White? Measuring the Myth of Colorblindness*, 7 J. EMPIRICAL LEGAL STUD. 886 (2010).

170. *Id.*

171. Brian A. Nosek & Mahzarin R. Banaji, *The Go/No-Go Association Task*, 19 SOC. COGNITION 625, 657 (2001).

172. *See* Justin D. Levinson, *Guilt by Implicit Racial Bias*, *supra* note 3.

from the test as constructed whether or not the implicit bias is actually driven by an association of black guilt or white innocence (or at least lack of guilt), or some combination of the two.

Fortunately, researchers have developed single category implicit measures that have allowed social scientists to conclude that some of the most common implicit biases are not only driven by out-group derogation, but also independently by in-group favoritism. Brian Nosek and his colleagues, for example, created the Go/No-Go Association Task (GNAT), a reaction time-based measure that requires participants to hit the space bar any time they see matched categories.[173] Thus, unlike a Race–Attitude IAT, a participant taking a Race–Attitude GNAT would have one racial category active (e.g., white) and would be asked to hit the space bar every time a positive (vs. negative) word appeared on a computer screen.[174] It is in this context that researchers have found that, when tested using the GNAT, not only do the most popular IATs represent implicit out-group derogation (e.g., an association between black and bad), but also a separate implicit in-group favoritism (e.g., an association between white and good).[175] Although not all IAT results have been tested using the single category GNAT format, these results indicate at least that the most commonly tested implicit racial attitudes indeed are marked by an implicit in-group bias for whites that is an entirely separate phenomenon than the implicit out-group derogation shown for blacks.

***

This Part has explained the social science underlying our contention that the implicit bias model of racial inequality in criminal justice must be broadened to include white favoritism. The following Part provides detailed examples of ways in which in-group favoritism may lead to unique benefits for white people in the criminal justice system.

III. IMPLICIT RACIAL BIAS AS WHITE FAVORITISM

Research on implicit black derogation has enriched and expanded the boundaries of traditional explanations for the persistence of racial disparities. Yet, if we want to provide a complete and behaviorally accurate account of the mechanics of race discrimination, we also must consider implicit bias as white favoritism. Virtually all of the black derogation narratives that legal scholars use to explain racial disparities in the criminal

173. *See* Nosek & Banaji, *supra* note 171, at 633.
174. *Id.* at 632.
175. *Id*. at 659.

justice system can be made richer by including a description of how implicit favoritism operates.

This Part focuses on how implicit white favoritism influences the decisions of the same actors that we highlighted in Part I: police officers;[176] jurors;[177] and legal professionals, including legislators, judges, prosecutors, and public defenders.[178] Throughout this Part, we take special care to highlight where favoritism is able to help explain disparities that are left unexplained by the black derogation model. Conversely, we point out a few places where favoritism adds relatively little to black derogation-based explanations. Mostly, though, derogation and favoritism operate in tandem. This dual-engine bias helps to explain why, for example, racial disparities tend to be the most dramatic in black defendant/white victim cases. Finally, throughout Part III, we highlight potential remedies and note whether the proposed solution would address favoritism alone or both white favoritism and black derogation.

A. White Favoritism and Police Officers

Implicit white favoritism, like implicit black derogation, can influence the range of discretionary judgments that police officers make. We focus on two examples that should be familiar from Part I: First, we explore favoritism in the decision to make a stop.[179] Here, though, we shift from citizens on foot to vehicular stops. Second, we question whether it is an inherent good that police officers exhibit slower reaction speeds in their decisions to shoot armed white suspects.

1. Traffic Stops

One of the best examples of how white favoritism can augment and amplify implicit bias-based accounts of police behavior is in the context of traffic stops. As in the stop-and-frisk context, data on vehicular stops indicates significant racial disparities in who the police stop and whether those stops convert into searches or arrests.[180] These disparities are so

176. *See supra* Part I(B) and accompanying text (describing implicit bias as out-group derogation in the context of police officers).

177. *See supra* Part I(C) and accompanying text (describing implicit bias as out-group derogation in the context of jurors).

178. *See supra* Part I(D) and accompanying text (describing implicit bias as out-group derogation in the context of legal professionals).

179. *See supra* notes 36–39 and accompanying text (describing how implicit black derogation can operate in the context of the decision to stop and frisk an individual suspect).

180. In states ranging from Missouri to North Carolina to Wisconsin, data from hundreds of thousands of traffic stops demonstrates that police disproportionately stop vehicles when black Americans are driving them. Matthew R. Durose et al., U.S. DEP'T JUSTICE, CONTACTS BETWEEN

culturally enmeshed, as is the sense that race and not the underlying traffic offense motivates the stops, as to have spawned a colloquial phrase used to describe vehicle stops involving black drivers: driving while black.[181]

Legal scholars have noted that the "driving while black" phenomenon privileges white Americans in that white people are less "vulnerable to being stopped by the police," less likely to "feel apprehensive about such encounters," and more likely to "feel empowered to exercise their constitutional rights."[182] This framing conceptualizes white privilege as the difference between *driving* and *driving while black*. In other words, it frames white privilege as the absence of black derogation. White favoritism, however, recognizes a separate distinction: the difference between *driving* and *driving while white*. Recall the study where after being primed with crime-relevant words, police officers more rapidly located a dot-probe on a computer screen when it appeared where a black face had been located.[183] A different variation of the same study demonstrated that police officers were slower to locate the dot-probe on the side of the screen that displayed a white face when they had been primed with crime-relevant words than when they had received no prime.[184] Thus, the study produced results suggesting that the concept of crime is *disassociated* with white faces.[185]

The idea that white Americans are disassociated with crime lends support to our point that there is a difference between driving and driving while white and that the latter phenomenon contributes to racial disparities in traffic stops. Imagine a car with a broken taillight or an expired license plate cruising down the road. As the car passes a police officer, the officer can either stop the car and ticket its driver or the officer can let the car go. Police officers charged with drug interdiction might use these pretextual circumstances to pull over the vehicle. Since police officers disassociate

---

POLICE AND THE PUBLIC 2005, at 1 (2007), *available at* http://bjs.ojp.usdoj.gov/content/pub/pdf/cpp05.pdf; Ian A. Mance, *Racial Profiling in North Carolina: Racial Disparities in Traffic Stops, 2000 to 2011*, TRIAL BRIEFS, June 2012, at 23–27 (2012) (describing results from a decade-long study of nearly 14 million traffic stops in North Carolina and noting that not only do the "numbers show black drivers are more likely to be stopped by police than white drivers, they show significant disparities in treatment once these motorists are in police control") (emphasis omitted).

181. The Urban Dictionary defines "Driving While Black" as "refer[ring] to the idea that a motorist can be pulled over by a police officer simply because he or she is black and then charged with a trivial or perhaps non-existent offense." *Driving While Black*, URBAN DICTIONARY, http://www.urbandictionary.com/define.php?term=driving%20while%20black (last visited Aug. 1, 2013); *see, e.g.*, David A. Harris, *"Driving While Black" and All Other Traffic Offenses: The Supreme Court and Pretextual Traffic Stops*, 87 J. CRIM. L. & CRIMINOLOGY 544 (1997).

182. Carbado, *supra* note 2, at 1002 (discussing this dynamic in the broader context of disparate police stops).

183. *See supra* note 35 and accompanying text.

184. *See supra* note 35 and accompanying text.

185. *See supra* note 35 and accompanying text.

white faces with criminality, a police officer might be more likely to let the broken taillight or expired plate "slide" if the driver is white than if the officer had not noticed the race of the driver as the car sped past. In the real world, then, disparities in traffic stops result both from negative-stereotype-driven pretextual stops of black drivers and by providing favoritism to white drivers in the form of subconsciously disassociating them with the concept of criminality.

Though favoritism and derogation operate in distinct fashions, the traffic stop context is one where similar remedies are likely to mitigate both concerns simultaneously. Unfortunately, Fourth Amendment doctrinal reform is unlikely to be one of the plausible remedial options. In *Whren v. United States*,[186] the Court rebuffed an opportunity to squarely confront racialized vehicle stops.[187] Whren had argued that because "the use of automobiles is so heavily and minutely regulated . . . a police officer will almost invariably be able to catch any given motorist in a technical violation," which permits officers to conduct stops based on "the race of the car's occupants."[188] The *Whren* Court, however, held that any subjective motivations that cause a police officer to stop a vehicle are irrelevant for Fourth Amendment purposes, even if those subjective motivations are grounded in race discrimination.[189] The Court then affirmed that the reasonableness requirement is satisfied so long as an officer has probable cause to stop the vehicle.[190]

Given the obstacles of doctrinal reform, legal scholars have proposed that police departments tie promotions and raises to so-called "hit rates," which refer to the percentage—rather than the absolute number—of times

186. Whren v. United States, 517 U.S. 806 (1996). On the relationship between *Whren* and racial disparities in traffic stops, see, for example, Mario L. Barnes & Robert S. Chang, *Analyzing Stops, Citations, and Searches in Washington and Beyond*, 35 SEATTLE U. L. REV. 673, 682 (2012) (describing racial disparities in traffic stops across several different jurisdictions and describing *Whren* as "unhelpful" for remedying these disparities because it "essentially empowers police to profile based on race just as long as officers can articulate any nominal violation of the law"). *See also* Kevin R. Johnson, *How Racial Profiling in America Became the Law of the Land*: United States v. Brignoni-Ponce *and* Whren v. United States *and the Need for Truly Rebellious Lawyering*, 98 GEO. L.J. 1005, 1007 (2010) (asserting that *Whren* "made legal challenges to profiling more, not less, difficult, thereby implicitly encouraging police officers to rely on racial profiles in law enforcement") (emphasis omitted).

187. *Whren*, 517 U.S. at 808.

188. *Id.* at 810.

189. *Id.* at 813 ("We of course agree with petitioners that the Constitution prohibits selective enforcement of the law based on considerations such as race. But the constitutional basis for objecting to intentionally discriminatory application of laws is the Equal Protection Clause, not the Fourth Amendment. Subjective intentions play no role in ordinary, probable-cause Fourth Amendment analysis.").

190. *Id.* at 819 (explaining that "the District Court found that the officers had probable cause to believe that petitioners had violated the traffic code" and holding that such a finding "rendered the stop reasonable under the Fourth Amendment").

that a vehicle search resulted in the discovery of contraband.[191] The focus on the search as opposed to the stop is necessary because the officer mostly will have correctly identified a traffic citation before pulling the vehicle over. Nonetheless, focusing on the success rate of searches makes sense because if the officer is using the traffic stop as a pretext for a more serious offense, then tying promotions to search hit rates should cause the officer to be more careful when deciding which vehicles to stop in the first place. The focus on hit rates, in turn, incentivizes officers to use non-racial criteria more highly correlated with the target crime.[192] Importantly, though, officers need to be trained both to not rely on blackness as confirming (e.g., the need to keep looking for objective factors) and to not rely on whiteness as disconfirming (e.g., the need to start looking for objective factors). Ironically, tying promotions and raises to hit rates might make some police officers more likely to search a vehicle with a white driver. In other words, it might be—though we do not know for sure—that officers are too solicitous of white drivers and thus they miss objective signs that the driver is engaged in the target offense.

2. To Shoot or Not to Shoot

For a more emotionally powerful consideration of why implicit favoritism matters, we return to the split-second decision of whether to shoot (or not shoot) a suspect.[193] In *Tennessee v. Garner*,[194] the Court addressed the case of a police officer who had shot an unarmed teenager as

191. Professor Richardson defines a "hit rate," which is also referred to as "arrest efficiency" as "the rates at which the police find evidence of criminal activity when conducting a stop and frisk." Richardson, *Police Efficiency and the Fourth Amendment*, *supra* note 4, at 1145 (explaining that implicit negative stereotyping can influence police decision making and noting that "hit rates" tend to be lower when the police search black citizens than when they search white citizens) (citing Andrew Gelman et al., *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N 813, 821 (2007)). For an analogous argument that prosecutors should be rewarded and promoted, in part, based on the degree to which their initial charging decisions match the ultimate convictions that they obtain, see Tracey L. Meares, *Rewards for Good Behavior: Influencing Prosecutorial Discretion and Conduct with Financial Incentives*, 64 FORDHAM L. REV. 851, 873 (1995). *But see* Stephanos Bibas, *Prosecutorial Regulation Versus Prosecutorial Accountability*, 157 U. PA. L. REV. 959, 1014 (2009) (describing Meares' "proposed combating [of] prosecutorial overcharging by financially rewarding prosecutors whose initial charges closely match the charges of conviction" as "promising in theory" but "probably unworkable in practice").

192. Richardson, *Police Efficiency and the Fourth Amendment*, *supra* note 4, at 1169 ("If their hit rates matter, officers will be motivated to be more accurate before conducting a *Terry* stop. This will likely translate into them gathering more unambiguous information of criminality than they currently obtain before conducting a stop. Thus, to the extent that a focus on hit rates changes officer incentives and motivates them to individuate, it has the potential to reduce the effects of implicit biases on officer behavior.").

193. *See supra* notes 43–48 and accompanying text.

194. Tennessee v. Garner, 471 U.S. 1 (1985).

the teen burglary suspect climbed over a fence in an attempt to flee arrest.[195] The officer acted under the color of Tennessee law, which permitted him the use of "all the necessary means to effect the arrest" in cases where the defendant, "after notice of the intention to arrest," proceeds to "flee or forcibly resist."[196] The Court invalidated the Tennessee statute, holding the use of deadly force is reasonable for Fourth Amendment purposes only where "the officer has probable cause to believe that the suspect poses a significant threat of death or serious physical injury to the officer or others."[197]

Though the *Garner* Court narrowed the scope of the permissible use of deadly force considerably, the "probable cause to believe that the suspect poses a significant threat of death or serious physical injury" standard leaves a lot of room for officer discretion.[198] Implicit racial bias thrives under such circumstances. To gauge whether implicit racial bias influences the decision to shoot, Professors Payne and Correll used the shooter bias paradigm discussed in Part I.[199] The immediate impetus for conducting the shooter bias studies stemmed from a 1999 police-involved shooting where New York City police officers fired forty-one rounds into Amadou Diallo, an unarmed twenty-two-year-old West African immigrant with no criminal record.[200] Recall that these shooter bias studies reveal that police officers shoot more rapidly when a black person is holding the gun.[201] Scholars treat this time-to-shoot disparity as evidence that police officers are not solicitous enough of black lives.

The fact that police officers shoot black citizens with guns faster than white citizens with guns does not necessarily mean that officers shoot blacks too fast and whites too slow. There is at least some indirect evidence that officers might sometimes wait too long to use deadly force against an armed white suspect. First, in the shooter bias studies, police officers shoot more slowly when the armed suspect is white.[202] Moreover, in a recent priming study, participants primed with a white face needed more frames to identify a degraded image of a weapon as it came into focus than when they were primed with a black face or when they did not receive a prime.[203] Thus, it could be that the implicit disassociation between white Americans

195. *Id.* at 3 (noting that the officer "suspected" that the suspect was unarmed, but fired his weapon regardless).

196. *Id.* at 4.

197. *Id.* at 3.

198. *Id.*

199. *See supra* notes 43–48 and accompanying text.

200. Correll et al., *Across the Thin Blue Line*, *supra* note 49; Correll et al., *The Police Officer's Dilemma*, *supra* note 43.

201. *See supra* notes 43–48 and accompanying text.

202. Eberhardt et al., *Seeing Black*, *supra* note 27.

203. *Id.*

and weapons leads officers to respond too slowly in the face of imminent danger.

Again, the remedy here—as in the traffic stop context—is to train police officers to focus more intensively on objective factors that correlate more closely with imminent dangerousness.[204] For instance, an FBI study of felonious assaults on police officers noted several early warning signs of weapons use including the number of times the individual touches what the officers believe to be a weapon, whether the offender assumes a "blading" position so as to conceal the side of his body where the weapon is hidden, and whether the offender is wearing multiple layers of clothing despite being in a warmer climate.[205] The strength of these warning signs as predictors of whether the suspect will use lethal force is beyond the scope of this Article. Our point is simply that relying on some set of objective criteria is likely to reduce emphasis on the whiteness of the suspect as a factor that inhibits the decision to shoot.

B. White Favoritism and Juries

Black derogation and white favoritism are distinct, though related, obstacles to an impartial jury. This Subpart focuses on the white favoritism side of implicit racial bias. It explores how white favoritism could operate in the context of two different decisions entrusted to American juries. First, we examine whether white favoritism influences the sentencing judgments of juries in life without possibility of parole cases. Second, we question whether white favoritism influences how jurors process and evaluate the evidence presented at trial; specifically, we focus on favoritism in the way jurors perceive the facts necessary to establish a self-defense claim.

1. Juvenile Life Without Parole

One of the gravest determinations that a jury can make is deciding whether to sentence a juvenile homicide offender to life without the possibility of parole.[206] To ensure that the jury has the information it needs to render this solemn judgment, the Supreme Court held in *Miller v. Alabama* that jurors must consider the background and characteristics of

204. Anthony J. Pinizzotto, Edward F. Davis & Charles E. Miller, U.S. DEP'T OF JUSTICE, FED. BUREAU OF INVESTIGATION, VIOLENT ENCOUNTERS: A STUDY OF FELONIOUS ASSAULTS ON OUR NATION'S LAW ENFORCEMENT OFFICERS (2006), *available at* http://freedomsadvocate.com/files/ref/FBi-ViolentEncounters-AStudyofFeloniousAssaultsonAmericasLawEnforcementOfficers-2003.pdf (last visited Feb. 20, 2014).

205. *Id.* at 56–57 (describing behavioral characteristics of suspects who might commit felonious assaults on police officers).

206. Miller v. Alabama, 132 S. Ct. 2455, 2469 (2012) (neglecting to hold that life without the possibility of parole is a per se cruel and unusual punishment for juveniles who commit murder).

the offender to determine whether the particular defendant is "the juvenile offender whose crime reflects unfortunate yet transient immaturity, [or] the rare juvenile offender whose crime reflects irreparable corruption."[207] In essence, then, jurors are asked to consider whether the fourteen-year-old is someone who is capable of redemption. This inquiry often centers on a prediction about whether the juvenile will mature into a responsible adult.[208]

As the Court explained in *Miller*, juvenile offenders have a "heightened capacity for change" because their "characters are not as well formed," they have an "underdeveloped sense of responsibility," and they tend to be "more vulnerable or susceptible to negative influences and outside pressures, including peer pressure."[209] These are transient traits for most juveniles, so the difficulty lies in sorting out those juveniles whom the jury believes will not change with age. How can implicit white favoritism shape resolution of these questions? Recall the Ultimate Attribution Error, which is the tendency to attribute the positive behaviors of in-group members as dispositional and negative behaviors as situational.[210] If a fifteen-year-old white male kills a store clerk while attempting to steal beer from the liquor store, the Ultimate Attribution Error would suggest that white jurors are more likely to deem the white juvenile's actions as "unfortunate" and attributable to the "transient" nature of youth; whereas the same white jurors could find that a fifteen-year-old black male defendant who committed the same violent crime did so because it reflects his "irreparably corrupt" disposition.[211] The white juvenile defendant is not simply benefiting from the absence of negative stereotyping. He receives a boost in that his negative actions are interpreted through a lens that tends to emphasize the role that his situation played in the offense—the natural inclination is to find him redeemable.

Such in-group favoritism attribution errors are, of course, not limited to the homicide context. Imagine a fourteen-year-old boy who committed a string of armed robberies. The defense asks the jury to consider the fact that his client has been addicted to "pills" since age twelve. Like juvenile status, addiction is strongly associated with "[i]mpairment in behavioral control," "[d]iminished recognition of significant problems with one's

207. *Id.* (quoting Roper v. Simmons, 573 U.S. 551, 573 (2005) (internal quotation marks omitted)). For recent scholarship on the import of *Miller*, see, for example, Carol S. Steiker & Jordan Steiker, Miller v. Alabama*: Is Death (Still) Different?*, 11 OHIO ST. J. CRIM. L. 37 (2013) and Barbara Fedders, *Defining the Role of Counsel in the Sentencing Phase of a Juvenile Delinquency Case*, 32 CHILD. LEGAL RTS. J., Fall 2012, at 25.

208. *Miller*, 132 S. Ct. at 2465 (explaining why "an irrevocable judgment about [the juvenile's] value and place in society" is "at odds with a child's capacity for change").

209. *Id.* at 2469, 2475.

210. Hewstone, *supra* note 13, at 311–12.

211. *See Miller*, 132 S. Ct. at 2469.

behaviors and interpersonal relationships," and impaired "perception, learning, impulse control, compulsivity, and judgment."[212] The teenager told the jury that he committed the robberies in order to get enough money to score more drugs because his mom and dad did not want him to have an afterschool job. Is this fourteen-year-old teenager a rotten apple, or is his addiction driving the criminal behavior? Race clearly should be an irrelevant consideration. Yet, the Ultimate Attribution Error suggests that jurors will be more likely to attribute the teenager's bad deeds to situational factors—i.e., his addiction—and not to dispositional factors—i.e., this teen is a born criminal—if the teen is white.[213] The Court held in *Graham v. Florida* that a juvenile could not receive a sentence of life without the possibility parole for a non-homicide offense, and thus, the sentencing decision in this hypothetical would fall to a judge in most jurisdictions.[214] On the other hand, addiction is a common theme in death penalty trials, too, and jurors must make similar assessments in those cases about the capacity for redemption in adult homicide offenders.[215] Thus, although powerfully illustrated within the juvenile context, there is no reason to believe that the same mechanisms for facilitating white favoritism do not also apply in the capital punishment context.

When considering potential remedies for white favoritism in the juvenile life without parole context, it is helpful to remember the baseline *Miller* assumption is that the vast majority of the juveniles who commit murder should not receive a life without parole sentence.[216] This baseline suggests that attempting to neutralize the interpretative "benefit of the doubt" that white defendants receive is not the best strategy. Instead, reformers might consider how to increase opportunities for jurors to empathize with the black juvenile. In individual cases, for example, lawyers representing a black defendant will need to pay special attention to emphasizing the situational aspects of the offense, including the transient nature of the behavioral and cognitive deficits association with youth, and explaining to the jury why that means that their client is not an irredeemably bad apple. White defendants benefit from this explicit

212. *Public Policy Statement: Definition of Addiction: Long Definition of Addiction*, AM. SOC'Y OF ADDICTION MED., http://www.asam.org/for-the-public/definition-of-addiction (emphasis omitted) (last visited Jan. 16, 2014).

213. Hewstone, *supra* note 13, at 315.

214. Graham v. Florida, 560 U.S. 48, 62 (2010) (holding that the Eighth Amendment bars life without the possibility of parole sentences for juvenile offenders who commit non-homicide offenses).

215. Robert J. Smith et al., *The Failure of Mitigation?*, 65 HASTINGS L.J. 1221, 1245 (2014) (listing recently executed offenders whose mitigation histories suggest chronic drug addiction).

216. *Miller*, 132 S. Ct. at 2469 ("[G]iven all we have said in *Roper*, *Graham*, and this decision about children's diminished culpability and heightened capacity for change, we think appropriate occasions for sentencing juveniles to this harshest possible penalty will be uncommon.").

framing, too, but they also are more likely to get the initial boost from the operation of the Ultimate Attribution Error.

Miller itself helped increase opportunities for the jury to empathize with a black juvenile defendant. The Court could have held simply that jurors must be given the option of declining to impose a life without parole sentence for a juvenile homicide offender. Instead, the Court imported a requirement from its death penalty jurisprudence that ensures that jurors engage in individual sentencing.[217] If the death penalty corollary is any indication, then this right to present mitigation evidence triggers the secondary right of reframing the assessment of effective assistance of counsel to include the mitigation investigation.[218] In other words, *Miller* is a white-favoritism-busting opinion in that it requires lawyers to investigate and present—and jurors to consider—the type of mitigating information about the black defendant's character background that the jurors could take for granted when the defendant is white.

2. Interpreting Evidence and Evaluating Witness Credibility

Though determining whether a juvenile homicide offender should receive a life without parole sentence is among the most serious responsibilities entrusted to juries, determining sentencing outcomes is not among the most frequent roles that jurors occupy. Fact finding is at the core of the jury function, however. Whether it's a low-level drug distribution case or a quadruple homicide, jurors must process, interpret, and evaluate evidence and also assess the credibility of witnesses. This Subpart highlights how implicit white favoritism can taint these core jury functions. It does so by focusing on several variations of a self-defense claim to illustrate how a single determination can elicit multiple openings for the corrosive effects of white favoritism.

A self-defense claim involving a white defendant and a black victim is a classic example of the interdependence of the two distinct, yet related, concepts of black derogation and white favoritism. In Part I, we explained that implicit negative stereotyping could facilitate a jury finding that a white defendant had a "reasonable" basis to believe that the black victim posed an imminent danger of serious bodily harm.[219] Remaining with the same hypothetical case of the white defendant and the black victim,

217. *Id.* at 2475 ("*Graham*, *Roper*, and our individualized sentencing decisions make clear that a judge or jury must have the opportunity to consider mitigating circumstances before imposing the harshest possible penalty for juveniles.").

218. *See, e.g.*, Rompilla v. Beard, 545 U.S. 374, 393 (2005) (reversing a death sentence and remanding for a new sentencing hearing where Rompilla's trial counsel provided prejudicially deficient representation in the penalty phase of Rompilla's capital trial).

219. *See supra* Part I(C) and accompanying text.

consider how white favoritism can operate as an independent side effect of black derogation. Recall the research on stereotype lift, which describes a performance-enhancing lift that occurs when a relevant out-group negative stereotype is activated. We described in Part I how negative stereotypes of black Americans are activated in the self-defense context: black Americans are implicitly associated with weapons and stereotyped as hostile, violent and prone to criminality. The simple, yet powerful, idea beyond stereotype lift, then, is that when the white defendant testifies at his own trial his awareness of those negative black stereotypes can enhance his own performance (and association with opposite, positive stereotypes) as he recounts the details of the altercation for the jury. Though the precise mechanisms behind this phenomenon are not yet well understood, it is at least easy to imagine the anxiety reduction that accompanies knowing your version of events is one that fits into a well-worn narrative: that your story is consistent with the automatic associations of white jurors and not against them.

In the previous variation of a self-defense claim, implicit favoritism served as a further insult to the injury caused by the implicit negative stereotyping of the black defendant. Now consider a self-defense claim raised by a black defendant in a case where the deceased was a white male. Negative stereotyping could influence how jurors decide, for instance, which party was the aggressor: the fact that black Americans are stereotyped as violent and hostile, for instance, could influence whether the jury believes the black defendant when he testifies that the deceased white male started the altercation. Yet, even if the risk of negative stereotyping evaporated, white favoritism could still facilitate racialized decision making. First, white Americans are stereotyped as peaceful and law-abiding.[220] When jurors filter the details of the altercation through a lens shaped by the association of these positive stereotypes with the deceased white person, it provides a boost to the prosecution's narrative that it was the black defendant—and not the dead white guy—that started the altercation.

Or consider a close case where jurors are grappling with the question of whether the black defendant used deadly force pursuant to a reasonable belief that the now-deceased white male posed an imminent and lethal threat. Recall that it took participants primed with a white face significantly more frames to recognize a degraded object as a weapon than the participants who did not receive a prime.[221] That the jurors might implicitly disassociate the deceased white male with a weapon could undercut a claim

220. *See* Levinson, *Forgotten Racial Equality*, *supra* note 3, at 398–99; *see also* Eberhardt et al., *Seeing Black*, *supra* note 27, at 181–83.

221. *See supra* notes 40–42 and accompanying text.

by the black defendant that the white male, for example, reached into his own pocket and retrieved a shiny object that appeared to the black defendant to be a gun. Considered alone, then, white favoritism is a powerful tool for skewing how jurors process and evaluate evidence and assess the credibility of witnesses (including, in the context of a self-defense claim, the defendant). Yet, in a case with a dead white person and a black defendant, white favoritism and black derogation operate together to create a doubly dangerous risk that race influences the decision of the jury.

C. White Favoritism and Legal Professionals

This Part has considered how white favoritism influences the discretionary judgments of police officers and jurors. This final Subpart uncovers the corrosive effect of white favoritism on legal professionals: the legislators who define crimes and the judges, prosecutors, and public defenders whom preside over, prosecute and defend criminal cases.

1. Legislative Judgments

America panicked in the 1980s as a crack cocaine "epidemic" spread throughout its urban core; the media clung to images of "crack murders" and fears that "crack babies" might eventually be part of a "'biological underclass' suffering from, among other things, permanent mental retardation, deviance, and an inability to perform basic self-care tasks."[222] Congress reacted with a wildly punitive law that treated crack cocaine as if it were one hundred times worse than cocaine in its powdered form.[223] The 100:1—now 18:1—sentencing disparity between crack cocaine and powder cocaine has become the rallying cry of those that believe that the war on drugs is a "war on blacks." As we explained in Part I, Professor Ogletree and his colleagues suggested a more nuanced understanding—that implicit negative stereotyping of black Americans as dangerous and prone to criminality likely fueled concern over crack murders and helped legislators to see the "proper" response to the crack epidemic as one involving increasingly punitive sanctions.[224]

Between 2005 and 2006, media outlets ran over one hundred stories about a new drug "epidemic."[225] Rather than crack, this "epidemic" centered on methamphetamine. Legislatures across the country soon

222. Ahrens, *supra* note 66, at 854.
223. *Id.* at 856–57 (detailing the 100:1 [now 18:1] crack–powder disparity).
224. Ogletree et al., *Coloring Punishment*, *supra* note 69, at 51–52.
225. Ahrens, *supra* note 66, at 862.

enacted new laws to address methamphetamine use.[226] Based on the legislative response to crack use, one might have expected a draconian legislative response. Yet, unlike with crack use, state legislatures' responses to the perceived methamphetamine epidemic focused more heavily on non-punitive measures such as prevention (e.g., restricting sales of the main ingredient), education, and treatment.[227] Two states—Illinois and Montana—even created separate rehabilitation-focused prisons for methamphetamine users and other states created separate rehabilitative-focused units in general population prisons.[228]

Thus, unlike crack, many legislatures did not respond to methamphetamine usage with draconian prison sentences. Instead, the general response from public officials was that "meth" use presented a public health problem that warranted a "sympathetic" response—one driven more by "pity" than fear.[229] The reasons for differences between the responses to crack and methamphetamine are difficult to pin down. Unlike crack,[230] methamphetamine is perceived as a "white drug,"[231] which suggests the possibility that the difference in racial association between crack and meth played a role in how people perceived the proper response to the respective drugs. Media reports on the meth epidemic focused on rural, poor, white people as meth users.[232] Public officials—most of whom are also white—did not "think crack and then see black," as Professor Ogletree and colleagues put it, but rather, as Professor Ahrens wrote, "perhaps the fact that the majority of methamphetamine users are white . . . changed the incentives for press, police, and politicians."[233]

---

226. *See e.g.*, FLA. STAT. § 893.1495 (2005); IND. CODE § 35-48-4-14.7 (2005); MISS. CODE ANN. § 41-29-315(2)(c)(ii) (2005); MO. REV. STAT. § 195.417 (2005); MONT. CODE ANN. § 50-32-502 (2005).

227. Ahrens, *supra* note 66, at 866. *But see* United States v. Hayes, 948 F. Supp. 2d 1009, 1027–29 (N.D. Iowa 2013) (noting that the federal sentencing guidelines treat methamphetamine harshly).

228. Ahrens, *supra* note 66, at 876.

229. *Id.* at 885.

230. *See* Jeanine L. Skorinko & Barbara A. Spellman, *Stereotypic Crimes: How Group-Crime Associations Affect Memory and (Sometimes) Verdicts and Sentencing*, 8 VICTIMS & OFFENDERS 278, 288, 291 (2013) (finding that people associate crack with African-Americans).

231. Ahrens, *supra* note 66, at 885.

232. *Id.* at 862.

233. *Id.* at 845; *see also* Deborah Ahrens, *Drug Panics in the Twenty-First Century: Ecstasy, Prescription Drugs, and the Reframing of the War on Drugs*, 6 ALB. GOV'T L. REV. 397, 432 (2013) ("The media face of Ecstasy . . . generally has been that of white teenagers—affluent, suburban, and privileged. . . . The coverage of prescription drug abuse, as it is unfolding, similarly seems sympathetic—users are often portrayed as persons who originally were sick or injured and slipped into drug abuse, and the public image of prescription-drug abuse is white, middle-class, and, interestingly, generation-spanning."). Teasing out all of the factors that caused disparate reactions to crack cocaine and methamphetamine are beyond the scope of this Article. *Id.* at 432–33 (listing "changes in the constitutional law of sentencing," i.e., the federal sentencing guidelines are no longer mandatory, and a loss of "political will" due to "practical fiscal considerations" as two other partial explanations for disparate legislative and media responses to crack cocaine, methamphetamine and ecstasy). The goal

Implicit favoritism plausibly influenced legislative decisions to treat methamphetamine use more leniently than crack use. Though separated by class divisions, white methamphetamine users may have received the benefit of being perceived implicitly as in-group members—whites are "us" and blacks are "them." As Perdue and colleagues' studies demonstrated, the simple activation of these in-group categorizations can have important consequences.[234] Indeed, the powerful cognitive associations between the in-group and positive traits, and the out-group and negative traits, can help to explain how two new drugs took divergent paths of legislative discourse and lawmaking. In addition, categorizing the majority of "meth" users as "us" not only facilitates connections to positive versus negative traits, but, combined with the attribution errors we have already discussed, makes it easier to see their drug use as situational—a consequence of rural poverty, not a driver of it—as opposed to dispositional.

2. Judges

If the primary role of juries is to decide contested facts in criminal cases, then the primary role of trial judges in those cases is to determine an appropriate sentence. In this Subpart, we focus on two related factors that can contribute to a judge's decision to issue a more lenient punishment than she might decree for the typical person who commits the same offense: the expression of remorse and a statement of apology.

Judge Mark Bennett, who sits on the United States District Court for the District of Iowa, recently wrote on the importance of allocution—the statement of the defendant after he has been found guilty and before the judge determines the sentence.[235] In his essay, Judge Bennett explained that remorse and apology are critical aspects of a successful allocution.[236] He wrote: "Genuine remorse is essential to my consideration of a downward variance,"[237] which is a request from the defendant for a sentence below the minimum sentence for the applicable guideline range. An assessment of whether someone has "genuine remorse" is an uncertain enterprise—there is no blood test or mathematical formula to gauge the sincerity of an apology. Judge Bennett, however, writes that he does not worry much about being "conned" by an insincere apology because "[s]incerity—or

here is to describe a mechanism that helps to explain how race played a significant role in the divergent responses assuming that commentators are correct that race did indeed play a significant role.

234. *See* Perdue et al., *supra* note 112.

235. Mark W. Bennett, *Heartstrings or Heartburn: A Federal Judge's Musings on Defendants' Right and Rite of Allocution*, 35 CHAMPION 26, 26–29 (2011).

236. *Id.* at 27–28.

237. *Id.* at 28.

lack of it—is usually easy to spot" and the "anguish" that accompanies genuine remorse is "hard to fake" because "[o]ne can sense it."[238] He continues, "[a]s most of our mothers told us when we were young: 'It's not what you say but how you say it' that's often more important."[239]

Consider how implicit white favoritism can operate in an assessment of whether an apology is sincere or an expression of genuine remorse. White defendants are stereotyped as trustworthy.[240] Thus, being primed with the image of the white defendant while he gives the allocution could activate the association between white and trustworthiness, which, in turn, helps to shape whether the judge finds the defendant to be sincere in his apology. At the same time, enhanced in-group empathy could facilitate the ability of a white judge to more deeply "sense" the "anguish" that accompanies genuine remorse. Consider, though, how Judge Bennett helps to guard against the possibility that white favoritism influences his decision. He writes, for example, that the content of the apology is important.[241] Bennett wants defendants to explain why exactly they are remorseful—to express how the crime that they committed harmed the victims.[242] Moreover, when considering the apology, the judge again pays close attention to the content, ensuring that at the very least the offender makes a clear apology to the victim as opposed to a general apology to society or the court.[243] We do not know that these objective criteria fully eliminate the risk of white favoritism, but they certainly mitigate it.

3. Prosecutors

Prosecutors exercise vast discretion in their charging and bargaining decisions.[244] Studies routinely find that prosecutors exercise this discretion in racially biased ways.[245] Consider the capital punishment context, and specifically, the related decisions of whether to charge a case capitally, and, if so, whether to consider a life without the possibility of parole sentence if the defendant is willing to plead guilty. Nearly three decades of robust social science research has established that offenders who kill white

238. *Id.*

239. *Id.* at 28–29.

240. Justin D. Levinson et al., *supra* note 7, at 556 n.210 (2014).

241. Bennett, *supra* note 235, at 28–29.

242. *Id.*

243. *Id.*

244. Bibas, *supra* note 191, at 960 (noting that prosecutors possess more unreviewable discretion than any other actor in the criminal justice system).

245. *See, e.g.*, Starr & Rehavi, *supra* note 75, at 3 (stating that most "sentence gaps can be explained by prosecutors' initial charging decisions").

victims are the most likely to receive a death sentence.[246] Recent research demonstrates that prosecutors play an outsized role in perpetuating this "race-of-victim" effect.[247] In this Subpart, we illustrate how implicit white favoritism could help to explain the persistence of these findings.

Consider a homicide case with a victim who is a white woman with stereotypically white features. The prosecutor must decide whether to seek the death penalty. There are two places where white favoritism can facilitate a choice to pursue a death sentence: during consultation with the victim's family and at the point where the prosecutor must decide whether the homicide itself is aggravated enough to warrant a death sentence.

Prosecutors often consult with the victim's family members before deciding on whether to pursue a death sentence.[248] Indeed, in the federal system, Department of Justice protocols require U.S. Attorneys to "consult with the family of the victim, if reasonably available, concerning the decision on whether to seek the death penalty" and to "include the views of the victim's family concerning the death penalty in any submission made to the Department."[249] As a prosecutor listens to the mother, sister, or spouse of the victim, she hears those family members recount their lost loved as someone who is smart, beautiful, kind, and—more than anything else—deeply missed. These are heart-wrenching details regardless of who the particular victim is or how much the prosecutor relates to the family. Yet, when the victim and prosecutor are both white, there is an opportunity for enhanced in-group empathy. Recall the research on the activation of the "pain matrix" region of the brain, which found that when witnessing an in-group member receive a painful injection, cognitive activity occurred for in-group, but not out-group, members.[250] Similarly, it might be easier for white prosecutors to imagine the white victim as a daughter, sister, or spouse as opposed to if the victim is a black woman with dark skin and Afrocentric features. Moreover, just as with the needle entering the cheek of an in-group member, a white prosecutor might quite literally feel the

246. *See, e.g.*, David C. Baldus et al., *Evidence of Racial Discrimination in the Use of the Death Penalty: A Story from Southwest Arkansas (1990-2005) with Special Reference to the Case of Death Row Inmate Frank Williams, Jr.*, 76 TENN. L. REV. 555, 573 (2009); Glenn L. Pierce & Michael L. Radelet, *Death Sentencing in East Baton Rouge Parish, 1990–2008*, 71 LA. L. REV. 647 (2011) (finding significant race-of-the-victim effects in Baton Rouge, Louisiana).

247. *See* John Blume et al., *Explaining Death Row's Population and Racial Composition*, 1 J. EMPIRICAL LEGAL STUD. 165, 167 (2004) (noting disparities in "black defendant-white victim cases confirm[ing] the well-known race-of-victim effect").

248. Memorandum Regarding U.S. Attorneys' Manual Death Penalty Protocol Revisions from Paul J. McNulty, Deputy Attorney General, to Heads of Department Components, United States Attorneys, at 3 (June 25, 2007), *available at* http://www.justice.gov/dag/readingroom/dag-memo-06252007.pdf.

249. *Id.*

250. Xu et al., *supra* note 14, at 8525.

pain of the white victim's family members more than if the victim's family members were black.

In *Kennedy v. Louisiana*, the Court explained that "resort to the [death] penalty must be reserved for the worst of crimes and limited in its instances of application"[251] and commented how "[i]n most cases justice is not better served by terminating the life of the perpetrator."[252] Thus, a second location where white favoritism can seep into a prosecutor's decision to charge a case capitally is at the point when she is evaluating whether the crime itself is sufficiently egregious relative to other homicide offenses.

Consider a study by Levinson and colleagues in which they designed a "Value of Life" IAT that measured implicit associations between race and worth.[253] As hypothesized, Americans more quickly associated words that connote value (e.g., worth) with white and words that connote a lack of value (e.g., worthless) with black.[254] Implicitly associating white Americans with the concept of worth can skew prosecutors' judgments of how harmful they perceive a homicide to be relative to other homicides. Moreover, these racially skewed worth judgments can be intertwined with other mechanisms of implicit white favoritism. Most death penalty schemes attempt to narrow eligibility for capital punishment by requiring the prosecution to prove both that the defendant committed first-degree murder and also the presence of at least one aggravating factor.[255] For example, the prosecution might need to prove that the murder was particularly heinous, atrocious, or cruel.[256] This is another area where the empathy and perception of pain studies are relevant.[257] As in the context of assessing the desires of the victim's family, being able to identify with the pain that the victim suffered elevates the likelihood that the prosecutor will more fully comprehend and *feel* the harm that has been bestowed on the victim—to find that the suffering the white victim endured was particularly cruel. This harm evaluation, colored as it is by disparate magnitudes of empathy and worth assessment, can greatly influence whether a prosecutor decides to pursue capital charges and, if so, proceed to trial.

---

251. Kennedy v. Louisiana, 554 U.S. 407, 446–47 (2008).

252. *Id.* at 447.

253. Levinson et al., *supra* note 240, at 556.

254. *Id.* at 560.

255. *See, e.g.*, Ring v. Arizona, 536 U.S. 584, 593 (2002) ("The State's law authorizes the judge to sentence the defendant to death only if there is at least one aggravating circumstance and 'there are no mitigating circumstances sufficiently substantial to call for leniency.'" (quoting ARIZ. REV. STAT. ANN. § 13-703(F) (2001))).

256. *See* Schriro v. Summerlin, 542 U.S. 348, 361 (2004) (Breyer, J., dissenting) (noting that "[t]he leading single aggravator charged in Arizona . . . requires the factfinder to decide whether the crime was committed in an 'especially heinous, cruel, or depraved manner'" (quoting ARIZ. REV. STAT. ANN. § 13-703(F)(6) (2003))).

257. Xu et al., *supra* note 14, at 8525.

4. Public Defenders

Public defenders, too, can propagate favoritism-based inequality. In Part I, we discussed the argument made by Richardson and Goff that implicit negative stereotyping can impact how overburdened public defenders "triage" their cases.[258] One approach to "triaging" an overwhelming caseload is to devote disproportionate attention to cases in which the lawyer believes that the defendant might be factually innocent.[259] Richardson and Goff cautioned against using the defender's perception of innocence to prioritize cases because implicit negative stereotyping "thrives" under conditions where defenders are forced to act on "speculative hunches" since they do not have adequate time or information to sort cases by guilt or innocence.[260]

Implicit favoritism can also influence an assessment of guilt or innocence under conditions of constrained time and information. Consider a public defender trying to decide which of two cases to prioritize. Both cases involve assault charges. Case One involves a black victim and a white defendant. The black victim claims that the white defendant became visibly agitated during a heated conversation and then punched the victim in the side of the head. The white defendant claims that during the conversation the black "victim" became angry and threw a punch at him. The white defendant punched back in self-defense. There are no witnesses. Case Two is identical except that the defendant is black and the victim is white. Would the public defender rank the white defendant case as more plausibly an innocence case?

Recall that white Americans are dissociated with the concept of crime. White Americans also are stereotyped as law-abiding and peaceful. As the public defender sits in an interview room with his client, if the white defendant serves as a racial prime, it could allow the defense lawyer to more easily believe his client's story that the black victim threw the first punch. The white defendant also could benefit if his public defender is white. Research on implicit racial bias and determinations of trustworthiness indicate that white Americans implicitly trust other white Americans. Finally, the white defendant could benefit from stereotype boost, which describes how members of positively stereotyped groups can receive a performance enhancing boost when either their group membership or the stereotype itself is directly or indirectly primed. Stereotype boost could help to explain how a defendant's "innocence

258. *See supra* notes 89–91 and accompanying text (discussing the pain problems).

259. Robert P. Mosteller, *Why Defense Attorneys Cannot, but Do, Care About Innocence*, 50 SANTA CLARA L. REV. 1, 3–5 (2010).

260. *See supra* notes 89–91 and accompanying text.

performance" could be enhanced as his public defender interviews him about the facts of his case. The simple idea here is that when his white identity, which is stereotypically disassociated with violent crime, becomes activated (even indirectly), his words, language, actions, and believability could become automatically enhanced. This enhanced performance, in turn, could easily affect the way he—and the strength of his innocence claim—is perceived by his counsel.

CONCLUSION

Implicit racial bias is a tool for understanding racial disparities in the criminal justice system. Legal scholars capitalizing on the availability of this tool have mostly used it at half strength. This is not to imply that research on how implicit racial bias as out-group derogation has not been important. To the contrary, explanations for how implicit negative stereotyping of black Americans influences the way that legislators, police officers, jurors, and legal professionals make decisions have illustrated that racial factors are pervasive in criminal justice even if those who inject them do so unintentionally.

Yet, the picture created by the implicit racial bias as out-group derogation model is necessarily an incomplete one. To provide a more behaviorally rich portrait, as well as to contribute to understanding of racial disparities to which the out-group derogation model cannot speak, this Article provided the first systemic account of implicit favoritism and its role in perpetuating disparities in the criminal justice system. We explained the multiple ways in which implicit favoritism operates (e.g., through priming, through bias in explaining why an event occurred, through enhanced in-group empathy) to give the reader a sense that implicit favoritism is an umbrella term that applies to a variety of mechanisms—all of which operate without conscious intention. We then illustrated how implicit favoritism can operate in the criminal justice system by focusing on how it can impact the discretionary decisions of legislators, police officers, jurors, and legal professionals. This importation of implicit favoritism into the legal literature on racial disparities in criminal justice means that we now have a more behaviorally accurate understanding of how racial disparities are perpetuated, a better sense of the magnitude of the problem, and an understanding that scholars will need to address how we might curb favoritism—as well as derogation—if we hope to achieve a racially fair criminal justice system.