

Unconventional Methods for a Traditional Setting: The Use of Neurointerventions to

Reduce Implicit Racial Bias in the Courtroom

by

Natalie Salmanowitz

Program in Bioethics and Science Policy
Duke University

Date: _____

Approved:

Nita Farahany, Supervisor

Felipe De Brigard

Neil Vidmar

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Arts in the Program in
Bioethics and Science Policy in the Graduate School
of Duke University

2015

ABSTRACT

Unconventional Methods for a Traditional Setting: The Use of Neurointerventions to

Reduce Implicit Racial Bias in the Courtroom

by

Natalie Salmanowitz

Program in Bioethics and Science Policy
Duke University

Date: _____

Approved:

Nita Farahany, Supervisor

Felipe De Brigard

Neil Vidmar

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Arts in the Program in
Bioethics and Science Policy in the Graduate School of
Duke University

2015

Copyright by
Natalie Salmanowitz
2015

Abstract

The presumption of innocence and the right to a fair trial lie at the core of the United States justice system. While existing rules and practices serve to uphold these principles, the administration of justice is significantly compromised by a covert but influential factor: namely, implicit racial biases. These biases can lead to automatic associations between race and guilt, as well as impact the way in which judges and jurors interpret information throughout a trial. Despite the well-documented presence of implicit racial biases, few steps have been taken to ameliorate the problem in the courtroom setting. This paper suggests that neurointerventions, such as computerized brain-training tasks and noninvasive brain stimulation techniques, have the potential to provide promising mitigation strategies in the near future. Through analyzing the various ethical and legal considerations, this paper contends that the use of neurointerventions with judges would be both justifiable and morally obligatory should safe and effective means become available. A similar argument is put forth for jurors, albeit in a more theoretical light due to practical and logistical barriers. Given that implicit racial biases can seriously undermine the fairness of the justice system, this paper ultimately asserts that unconventional de-biasing methods warrant legitimate attention and consideration.

Contents

Abstract.....	iv
Acknowledgements.....	viii
Introduction.....	1
1. Defining the Problem: A Primer on Implicit Racial Biases in the Courtroom.....	4
1.1 Fairness, Impartiality, and The Nature of Biases.....	4
1.1.1 The Meaning of Impartial.....	4
1.1.2 Explicit Biases.....	6
1.1.3 Implicit Biases.....	6
1.2 Measuring Implicit Racial Biases.....	7
1.3 Implicit Racial Biases in the Courtroom.....	10
1.4 Exacerbating Factors.....	13
2. Proposed Strategies for Mitigation.....	15
2.1 Frequently Proposed Solutions.....	15
2.2 Criticisms of Proposed Solutions.....	17
2.2.1 Raising Awareness.....	18
2.2.2 Screening Methods.....	20
2.2.3 Weakening the Stereotypes.....	22
2.2.4 Diversification Efforts.....	23
2.3 An Alternative Proposal: Neurointerventions.....	25
3. Potential Neurointervention Techniques: A Proof of Concept.....	27
3.1 Overview of the Proposed Mechanisms.....	27

3.1.1 Computerized Brain-Training Tasks.....	27
3.1.2 Noninvasive Brain Stimulation.....	30
3.2 Relevant Safety and Practical Considerations	33
3.2.1 Computerized Brain-Training Tasks.....	34
3.2.2 Noninvasive Brain Stimulation.....	35
3.3 The Need for Tailored Research Efforts.....	37
3.4 Neurointerventions as a Feasible Prospect	39
4. Neurointerventions in Context: The Case of Judges.....	41
4.1 Neurointervention Policies: Painting the Picture.....	42
4.2 A Principle-Based Approach	42
4.2.1 Autonomy	43
4.2.2 Nonmaleficence	44
4.2.3 Beneficence.....	44
4.2.4 Justice.....	45
4.3 The Ethics of Voluntary Neurointerventions.....	45
4.3.1 Autonomy	45
4.3.2 Nonmaleficence	47
4.3.3 Beneficence.....	48
4.3.4 Justice.....	49
4.3.5 The Ethical Permissibility of Voluntary Neurointerventions	50
4.4 The Ethics of Pseudo-Compulsory Neurointerventions	50
4.4.1 Autonomy	50
4.4.2 Nonmaleficence	54

4.4.3 Beneficence.....	55
4.4.4 Justice.....	55
4.4.5 The Ethical Permissibility of Pseudo-Compulsory Neurointerventions	57
4.5 Neurointerventions as a Moral Obligation.....	58
4.6 Neurointerventions and Judges: A Worthwhile Consideration	61
5. Neurointerventions and Juries.....	63
5.1 Neurointervention Policies in Juries	64
5.2 The Ethics of Voluntary Neurointerventions.....	65
5.3 The Ethics of Pseudo-Compulsory Neurointerventions	67
5.3.1 Autonomy	68
5.3.2 Nonmaleficence	71
5.3.3 Beneficence.....	72
5.3.4 Justice.....	72
5.3.5 The Ethical Permissibility of Pseudo-Compulsory Neurointerventions	73
5.4 Neurointerventions as a Moral Obligation.....	76
5.5 Neurointerventions in Juries: Conclusions from a Conceptual Analysis.....	77
6. Neurointerventions in the Courtroom: Final Remarks and Future Directions.....	79
6.1 Summary.....	79
6.2 Future Directions	81
References.....	85

Acknowledgements

I would like to thank Professor Nita Farahany and Dr. Thomas Douglas for their guidance and feedback on this project.

Introduction

In 2011, the *Proceedings of the National Academy of Sciences* published a study that subsequently made international headlines. An analysis of parole decisions among eight judges in Israel revealed a surprising factor that influenced the outcome of a case. While one might have expected this factor to involve the gravity of the crime committed or the length of time served, it turned out to be something completely unrelated to the defendant or the crime: namely, the degree to which the judge was hungry or tired. Decisions to grant parole were highly correlated with when the case was heard in relation to a judge's break. Those heard right before a break were more likely to be denied, whereas those heard soon after were more likely to be granted (Danziger, Levav, & Avnaim-Pesso, 2011). The results of this study prompted significant uproar, as evidenced by headlines such as "When Lunch is Served, So is Justice" (Melnick, 2011) and "Hungry Judges Dispense Rough Justice" (Corbyn, 2011). The fact that a justice system could involve such patently extralegal factors seemed to threaten the court's legitimacy and spark doubt about whether justice was actually being delivered.

While this study is certainly interesting, its results are not nearly as groundbreaking or shocking as the media's response might suggest. The impact of extralegal factors in the courtroom is a well-documented phenomenon, and can influence both judges and jurors. In fact, whereas hunger and fatigue are sensations of which individuals are consciously aware, opinions and decisions can also be affected by factors unbeknownst to judges or jurors themselves. For example, the attractiveness, the demeanor, and the clothing of defendants can all play a role in our perceptions of their credibility and innocence (Bull,

1982; Greene & Heilbrun, 2011). Moreover, implicit reactions, such as racial biases, are not only widespread, but also can shape the evaluation and interpretation of information presented during a trial (Kang et al., 2012; Rachlinski, Johnson, Wistrich, & Guthrie, 2009). Implicit racial biases are perhaps even more unsettling than factors such as attractiveness, since while assessments of the latter can vary from one person to the next and might balance out across judges and jurors, the majority of the population has implicit racial biases that trend in a specific direction (Banaji, Bazerman, & Chugh, 2003; Mooney, 2014). The problem of extralegal factors in the courtroom therefore extends much deeper and wider than hungry or tired judges.

Given the amount of existing literature on extralegal factors, in particular implicit racial biases, what should warrant consternation is not the mere presence of such factors in the courtroom, but the relative lack of steps taken to mitigate them. While some countermeasures have been suggested, none appear to be sufficiently promising solutions. This paper aims to propose and examine an unconventional method to reduce implicit racial biases in the courtroom, specifically the use of neurointerventions. Although the actual materialization of such a prospect is still distant, researchers are beginning to discover various techniques that have the potential to ameliorate the impact of these biases.

This paper will argue that the use of neurointerventions in judges and jurors to reduce implicit racial biases should not be considered just ethically permissible, but morally required as well. The paper will approach these arguments in the following manner. First, a summary of the literature on implicit racial biases will be provided, with special

attention to studies in the courtroom setting. The second chapter will offer an overview and critique of commonly suggested countermeasures, while the third chapter will propose the use of neurointerventions as an alternative strategy. Fourth, the case for neurointerventions in judges will be made, addressing the ethical and legal considerations that might ensue. Lastly, this framework will be applied to the jury context, articulating the main divergences from the case of judges.

Given that specific neurointerventions have not yet been designed for use in the courtroom setting, this paper should be read primarily as a thought experiment. As a result, while certain practical considerations will be addressed, the ultimate purpose of this paper is to demonstrate the advantage of incorporating neurointerventions into the courtroom as opposed to delineating plans for their actual implementation. The following chapters will argue that while a neurointervention policy with jurors might remain in the theoretical realm given the unique and complex characteristics of the jury system, the use of neurointerventions in judges should be construed as a serious proposition (should the successful development of effective neurointerventions come to fruition). At the very least, this paper seeks to guide future discussion on novel and unconventional methods to reduce implicit racial bias in the courtroom.

1. Defining the Problem: A Primer on Implicit Racial Biases in the Courtroom

1.1 Fairness, Impartiality, and The Nature of Biases

1.1.1 The Meaning of Impartial

As articulated in the Bill of Rights, defendants are entitled to a fair and impartial trial (U.S. Const. amend. VI). For this reason, the justice system possesses mechanisms to screen for clear instances of bias in judges and jurors, such as motions for recusal and the voir dire process. In the case of judges, parties may request that the judge remove himself if there is reason to suspect any bias (e.g., 28 U.S.C. § 144). With jurors, the voir dire process enables attorneys and judges to ask probing questions of potential jurors in order to uncover possible signs of partiality (e.g., Fed. R. Crim. P. 24).

However, while this concept of impartiality permeates multiple legal standards, an official explanation of the term is lacking. As a result, it is unclear to what extent impartiality relates to the notion of fairness, as contemplated by due process, equal treatment under the law, and the assurance that each defendant receives the protections he is guaranteed (Congressional Research Service, 2014). Specifically, how impartial does a trial have to be in order to meet this description of fairness, and at what point does the presence of partiality violate such standards? While this chapter does not purport to definitively answer this question, it is necessary to provide a baseline understanding of impartiality in order to frame the arguments that this paper will put forth.

To start, in cases of perceived bias in judges or jurors, simply identifying a possible source of bias is generally insufficient grounds for disqualification; instead, there needs

to be a reasonable expectation that certain factors or characteristics would actually interfere with the judge or juror's ability to "be impartial" (Geyh, 2010; Howe, 1995). In fact, during voir dire, individuals expressing potential bias may still be selected for jury service if they assert that their biases can be put aside to "impartially" hear the case (Vidmar & Hans, 2007). Together, these facts suggest that impartiality does not involve complete neutrality or a lack of preexisting opinions on issues relevant to the trial. Instead, impartiality seems to entail the absence of preconceived notions that would *impact* a judge or juror's assessment of information presented throughout the case in a manner that *avored* or *disavored* a defendant. For instance, pretrial publicity can lead to a change in the trial venue on the basis that exposing judges or jurors to prejudicial information before the trial might impede their ability to (a) hear the case with an open mind and (b) adhere to the presumption of innocence; likewise, explicit conflicts of interest can disqualify judges and jurors for the same reasons (Congressional Research Service, 2014; Geyh, 2010). Moreover, federal regulations allow motions for recusal if the judge appears to "[have] a personal bias or prejudice either against [the party] or in favor of any adverse party" (28 U.S.C. § 144). This language supports the claim that influential factors providing advantages or disadvantages to a defendant are incompatible with the concept of impartiality.

The remainder of this chapter will argue that implicit racial biases render judges and jurors incapable of meeting this basic conceptualization of impartiality. Not only can these biases alter the way in which individuals interpret information, but they also tend to *disproportionately* disadvantage members of one race. The following section will explain

how general biases persist in the courtroom in the first place, and why implicit racial biases in particular pose such a threat to the fairness and impartiality of trials.

1.1.2 Explicit Biases

Despite protocols intended to screen for bias (such as motions for recusal and voir dire), partiality still plays a large role in the courtroom for three main reasons: the bias blindspot (Kang et al., 2012), the social desirability effect (Vidmar & Hans, 2007), and the existence of implicit biases (Larson, 2010; Reynolds, 2013; Roberts, 2012).

According to the bias blindspot, people are aware that biases occur in the general population, but tend to overestimate their own degree of impartiality. Next, the social desirability effect captures a perceived need for others to view us in an unbiased and socially respectable light. This effect might therefore prevent people from disclosing biases that are regarded as taboo. Collectively, the bias blindspot and the social desirability effect reduce the likelihood that judges and jurors will recognize their own biases and willingly admit to them. Lastly, motions for recusal and the voir dire process primarily target explicit biases, which are attitudes or prejudices that individuals *consciously* possess. As a result, the justice system leaves untouched many implicit biases that judges and jurors are not aware of having, but that are believed to account for a disturbingly high percentage of behavior (Rachlinski et al., 2009).

1.1.3 Implicit Biases

Implicit biases, while sometimes consistent with overt attitudes, are often completely separate from one's explicitly held beliefs (Banaji et al., 2003). In fact, these biases are the inevitable byproduct of having efficient brains. In order to quickly make sense of all

the stimuli bombarding us, we make generalizations about various individuals, social groups, and situations in a way that enables prediction of subsequent interactions and outcomes (Amodio, 2014). While the ability to make generalizations is frequently harmless and even quite useful in terms of efficiency, such judgments can also lead to negative assessments of individuals as well as stereotypical associations between specific attributes and identity groups (Ward, 2012). The automatic and influential nature of these biases can be appreciated by neuroscientific research involving implicitly prejudiced or stereotyped evaluations. For example, studies have explored activation in regions such as the amygdala, the anterior insula, and the anterior temporal lobe, which are implicated in threat processing, disgust reactions, and social stereotyping, respectively (Amodio, 2014). Researchers have observed increased activation in these three areas when white participants view black faces, and this heightened reactivity is positively correlated with the degree of implicit racial bias (Amodio, 2014). Accordingly, the fact that initial reactions of threat and disgust can arise when simply viewing a member of a racial outgroup presents serious problems in the trial context, where judges and jurors are expected to be open-minded upon entering the courtroom.

1.2 Measuring Implicit Racial Biases

Before delving into studies examining implicit racial biases in the courtroom, it is helpful to first explain how researchers identify the presence and magnitude of these biases, as well as describe the patterns that typically emerge. The most commonly used method is the Implicit Association Test (IAT), which measures reaction times during a sorting task (Nosek, Greenwald, & Banaji, 2005). The race version of the IAT shows

participants two paired words on both upper corners of a computer screen. A participant will either see a “stereotype-congruent” arrangement of Black/Bad in one corner and White/Good in the other, or a “stereotype-incongruent” arrangement with pairings of Black/Good and White/Bad. When presented with black and white faces, as well as positive and negative words, participants must match the particular stimulus with its appropriate label in one of the corners of the screen. After a participant completes trials with both stereotype-congruent and incongruent pairings, the average reaction times for the two trials are compared. For instance, the results for those who respond faster for stereotype-congruent pairings would suggest a preference for white faces, which can be further categorized according to a slight, moderate, or strong degree. The IAT can be used for many stereotype-based associations other than race (such as gender, age, etc.), and has been taken by over 4.5 million people. Interestingly, not only do 75% of all participants show a preference towards white faces (Banaji et al., 2003), but non-white participants tend to also exhibit a preference for white faces (while simultaneously displaying a greater preference than average for members of their own race (Project Implicit, 2011)). Additionally, researchers have modified the Race IAT to test the association between race and guilt. This version of the test has been coined the Guilt IAT, and has shown that individuals commonly associate black men with terminology related to criminal culpability (Levinson, Cai, & Young, 2010). These results might be especially concerning in light of confirmation biases, which hold that people are more likely to interpret evidence in a way that aligns with their presuppositions (Kahneman, 2011).

In addition to the IAT, two other tests are frequently used to measure implicit racial biases in the criminal context. The Weapons Identification Task, also a measure of response time, displays a picture of either a gun or a tool on the computer screen and asks participants to identify the object presented (Payne, 2001). However, prior to seeing the object, a white or black face is quickly flashed on the screen. Studies have revealed three particularly important findings. First, when white participants have to make these judgments in a constricted time frame, they are more likely to incorrectly identify a tool as a weapon when primed with a black face, as well as incorrectly identify a weapon as a tool when primed with a white face (Payne, 2001). Second, when given unlimited time to complete the task, participants are more accurate in their judgments, but correctly identify a weapon faster after seeing a black compared to a white face (Payne, 2001, 2006). Lastly, this pattern of results persists even when participants are explicitly warned to disregard the faces presented prior to the objects and to not let them skew subsequent judgments (Payne, Lambert, & Jacoby, 2002). A similar paradigm, the Shoot No-Shoot Test, involves a videogame in which white and black men appear on the screen, and either have a weapon or a neutral object in hand (Correll, Park, Judd, & Wittenbrink, 2002). Designed to mimic a situation often experienced by police officers, participants are told to shoot those holding a weapon while their response times and accuracy levels are measured. A study involving black and white participants found that both groups tended to shoot unarmed men more frequently when the target was black (as well as fail to shoot an armed man more often when the target was white). Importantly, this study also measured participants' knowledge and subjective evaluation of cultural stereotypes

associating black men with violent crime. While the degree to which participants *agreed* with this stereotype was unrelated to their performance on the task, knowledge of the stereotype was positively correlated with incorrect shooting responses (Correll et al., 2002). Thus, simply being aware of the stereotype influenced decisions, regardless of one's explicit attitudes.

1.3 Implicit Racial Biases in the Courtroom

Although the methods described above have frequently been used to examine implicit racial biases in the general population, a few studies have applied these concepts specifically to the courtroom setting. Given the difficulties of conducting real-time experiments during trials themselves, most studies addressing judges and jurors involve evaluations of hypothetical scenarios or post-hoc analyses of trial outcomes. Nevertheless, these studies reveal that implicit racial biases in judges and jurors are likely to have potent impacts on criminal justice proceedings.

To start, research has illuminated the effect of a defendant's race on judicial behavior in both hypothetical situations and actual sentencing patterns. In a study exploring implicit racial biases in judges, Rachlinski and colleagues (2009) administered the Race IAT to judges of varying races, genders, and jurisdictions. Overall, the white judges demonstrated a strong preference for white faces while the black judges did not display an overarching trend one way or the other. The researchers then provided these judges with a hypothetical scenario in which either the defendant was black and the victim was white, or vice versa. Interestingly, the authors observed significant correlations between IAT scores and conviction decisions only among the black judges. When the defendant

was white, black judges with an IAT preference for black faces were more likely to convict than were those with a white preference. This pattern flipped when the defendant was black, meaning that black judges with a preference for black faces were now less likely to convict. The researchers suggest that the lack of correlation between IAT scores and conviction decisions in white judges might stem from the fact that the majority of the judges were cognizant of the experiment's purpose; the behavior of the white judges might therefore be explained by social desirability effects, as they might have anticipated that the researchers were expecting them to respond in a certain way. Nonetheless, this study reveals that implicit racial biases are not only prevalent among judges, but that they also can occur in judges of differing races.

With respect to the impact of race on actual sentencing decisions, studies have found that on average, black defendants are given longer sentences than white defendants, and are also more likely to receive the death penalty (Kang et al., 2012; Rachlinski et al., 2009). In fact, a study exploring the correlation between race and sentence length in a Florida prison population found that within white and black subgroups, the more an inmate possessed Afrocentric physiognomic features, the longer his sentence (Blair, Judd, & Chapleau, 2004). These results have been explained by referencing the neuroscientific literature cited earlier; since the amygdala responds to threat, and black individuals are often associated with violent crime, this stereotype of "black-crime" might induce amygdala-driven threat responses that subconsciously guide judicial sentencing decisions (Papillon, 2013).

Turning to jurors, a study using a mock jury examined the relationship between implicit racial bias and the interpretation of evidence in a trial. The researchers manipulated one piece of evidence between the experimental and control group, namely whether the hand of a masked gunman in a picture was light or dark skinned. At the end of the trial, the majority of mock jurors were unable to report the skin tone of the person depicted in the image, yet on average, those who saw the dark skinned version deemed the defendant to be guiltier than those presented with the light skinned version (Levinson & Young, 2010). In a similar study, researchers incorporated both the Race IAT as well as the Guilt IAT. When presented with ambiguous evidence and asked to evaluate the degree to which it informed judgments of guilt or innocence, those with higher scores on both types of the IAT were more likely to rate such evidence as supporting a guilty verdict (Levinson et al., 2010). In other words, participants who possessed preferences for white faces and more frequently associated blacks with criminal guilt tended to view vague evidence as a testament to the defendant's culpability. Hence, even if implicit racial biases do not influence the ultimate decisions of mock jurors, at the very least they can impact how jurors weigh and interpret information provided during a trial. Additionally, in some mock jury studies, researchers have found what have been coined "race salience" effects, whereby white jurors are more likely to convict black defendants in hypothetical scenarios when the subject of race is not explicitly called to their attention, but are much less likely to convict when race plays a prominent role in the details of the case (Sommers & Ellsworth, 2001). The authors explain these findings through the lens of social desirability effects, in which white participants might be

making a purposeful effort to avoid displaying any signs of racism in situations where others might be expecting them to do so.

In sum, although there is limited research regarding the impact of implicit racial biases on actual case outcomes, the studies we do have suggest that judges and jurors are susceptible to these biases in a way that can influence evaluations throughout the trial process.

1.4 Exacerbating Factors

In addition to documenting the prevalence of implicit racial biases, researchers have also discovered certain factors that serve to exacerbate these biases. First, as noted earlier, stereotypes enable us to process information quickly and make automatic judgments. These fast-paced decisions, which can lead to the use of negative stereotypes, occur more frequently in situations involving vague information or general uncertainty (Reynolds, 2013). As Larson (2010) notes, trials are rife with uncertainty; after all, if a case were patently clear, it would most likely result in a plea agreement (or no charge being brought) and would not be tried in court. Second, as seen with social desirability and race salience effects, individuals do have the capacity to monitor the influence of implicit or explicit biases under certain circumstances (Rachlinski et al., 2009; Sommers & Ellsworth, 2001). However, this ability requires vigilant self-regulation (Banaji et al., 2003), and tasks that involve a high degree of cognitive effort will subsume a significant portion of one's mental control capacities. Thus, in situations (much like the courtroom setting) where individuals must pay close attention to specific facts and details, the ability to monitor biases will be substantially decreased (Papillon, 2013). Third, emotions are

known to compromise decision-making capacities and magnify the influence of implicit biases (Bublitz, 2015). In fact, simply eliciting feelings of disgust prior to taking an IAT can make preferences against an outgroup more severe (Dasgupta, Desteno, Williams, & Hunsinger, 2009). This finding could pose serious concerns in the courtroom setting if initial visceral reactions in response to the defendant's race or the nature of the crime serve to exacerbate the influence of implicit biases. The amplifying impact of emotions on implicit biases could also occur if judges or jurors become stressed during trial; when making evaluations regarding the morality of a situation, stress has been shown to decrease an individual's ability to comprehensively take relevant details into account as well as increase one's reliance on automatic judgments (Caviola & Faber, 2014). Therefore, not only are implicit racial biases a prominent factor during trials, but the courtroom setting itself can also magnify their influence.

2. Proposed Strategies for Mitigation

2.1 Frequently Proposed Solutions

Given the prevalence of implicit racial biases in the courtroom, many legal scholars and psychologists have suggested strategies to mitigate their presence and influence. These proposed interventions include (1) raising awareness, (2) screening with the IAT prior to trial, (3) weakening stereotypical associations, and (4) increasing diversity among judges and jurors. Each strategy will now be addressed in turn.

Arguments in favor of simply raising awareness draw support from literature on self-regulation and active monitoring. In particular, proponents frequently cite the social desirability and race salience effects discussed above to show that focusing attention on race can help motivate heightened scrutiny of one's own decisions (Rachlinski et al., 2009; Roberts, 2012). Similarly, scholars point to research suggesting a connection between motivation and behavior (Amodio, 2014; Ward, 2013), claiming that if courts make judges and jurors more aware of implicit racial biases, they will be more inclined to regulate their own behavior, which will in turn help to reduce the influence of such biases (Kang et al., 2012; Rachlinski et al., 2009; Reynolds, 2013; Roberts, 2012). In fact, Roberts (2012) proposes educating jurors about implicit racial biases through juror orientation videos, as these materials are designed to harness jurors' sense of civic duty and galvanize them to contribute their best efforts in the pursuit of justice.

In addition to raising awareness, some scholars endorse implementing screening protocols into the jury selection process in order to eliminate those with strong implicit racial biases. For example, Larson (2010) proposes administering the IAT to potential

jurors during voir dire. He contends that not only would such testing increase the motivation in all jurors to reduce their biases, but it would also flag individuals who possessed severe biases. This would subsequently facilitate the decision process for attorneys making challenges for cause, effectively removing these individuals from the jury pool for that trial. Such a screening strategy would similarly apply to judges, whereby IAT testing would take place before the assignment of a judge to a case.

Third, multiple scholars advocate interventions that target the stereotypes themselves (Kang et al., 2012; Rachlinski et al., 2009; Wittenbrink, Judd, & Park, 2001). By weakening the association between black men and violent crime, for example, the stereotype might become less hardwired and automatic. Proposed mitigation methods involve presenting people with counter-stereotypical examples, whether through placing portraits of famous black historical figures on the wall (Kang et al., 2012), or by having a black individual proctor an IAT (Banaji et al., 2003). Many proponents also cite a study by Dasgupta and Greenwald (2001) in which the presentation of both revered black leaders and disliked white historical figures reduced implicit racial bias scores. In addition to displaying counter-stereotypical images, researchers have also explored the use of mental imagery exercises to weaken negative stereotypes. In fact, mental imagery exercises, in which individuals visualize certain scenarios, have been shown to be very effective in helping people carry out goal-directed actions (Blair, Ma, & Lenton, 2001). Researchers attribute these results to the dynamic nature of imagination and the involvement of the same sensory activation patterns as those needed in real-life situations (Knäuper, Roseman, Johnson, & Krantz, 2009). In the context of implicit racial biases,

studies asking individuals to visualize positive counter-stereotypes have found considerable success in reducing IAT scores (Devine et al., 2012; Turner & Crisp, 2010).

The fourth and final commonly proposed intervention involves diversifying the pool of judges and jurors. Some scholars believe that the presence of judges with different viewpoints or identity-related characteristics might broaden the perspectives of their fellow colleagues (Rachlinski et al., 2009). With juries, not only do more diverse mock juries display heightened scrutiny and engage in more carefully reasoned discussions (Vidmar & Hans, 2007), but the anticipation of serving on a diverse mock jury can also reduce implicit racial biases (Sommers, 2006). For example, in a study comparing a homogenous white mock jury to a diverse mock jury, those on the diverse jury were significantly less likely to consider a black defendant guilty prior to the group discussion. The author hypothesizes that the element of race was made more salient for the white jurors on the diverse mock jury, who as a result made more concerted efforts to avoid exhibiting bias; Sommers (2006) additionally claims that this diverse panel composition impacted the way in which jurors interpreted the evidence presented to them. Furthermore, scholars suggest that enhancing the diversity of jurors would differentially increase the type or direction of implicit biases within the jury, which might serve to nullify at least some of these biases (Kang et al., 2012; Roberts, 2012).

2.2 Criticisms of Proposed Solutions

Although these interventions certainly have their merits, the following discussion will argue that all four methods are insufficient to substantially reduce implicit racial biases in the courtroom. This section does not intend to imply that these methods would not be

useful interventions in their own right, but instead maintains that they would not serve the purpose of reducing implicit racial biases to the degree that their proponents suggest.

2.2.1 Raising Awareness

Increasing the motivation to facilitate a fair trial and promoting awareness of implicit biases are both laudable goals for obvious reasons. However, in the context of race, three main concerns arise: (1) such an intervention might not have an appreciable effect in many individuals, (2) this method might actually have a counterproductive effect in some people, and (3) many of the factors known to exacerbate implicit racial biases are also those required for these interventions to succeed.

As previously mentioned, proponents of raising awareness frequently cite studies in which individuals demonstrate less biased behavior due to race salience and social desirability effects (Dasgupta & Greenwald, 2001; Rachlinski et al., 2009; Sommers & Ellsworth, 2001). However, while these studies do indeed exist, there is also a body of literature suggesting that simply being aware of bias does not do much to ameliorate its influence. For example, recall the study on the weapon bias effect, where individuals were primed with black or white faces prior to seeing a tool or a gun, and were explicitly told to avoid letting the face impact their decisions (Payne et al., 2002). Despite this warning, participants still tended to identify the object as a weapon when primed with a black face; in other words, calling attention to the risk of implicit racial bias had a negligible effect on reducing its impact. Moreover, in a longitudinal study involving various self-regulation strategies, researchers found that participants in the control group, who simply took the IAT and were told their results, did not show reduced bias on

subsequent IATs (Devine, Forscher, Austin, & Cox, 2012). While the control group in this experiment was not explicitly educated about the influence and nature of implicit racial biases after taking the IAT, these findings nevertheless question the assertion that basic awareness can result in a significant reduction of bias.

Solely raising awareness not only might be futile in many judges and jurors, but when awareness interventions take the form of emphasizing race salience or facilitating social desirability effects, the outcome might actually contribute to the impact of bias rather than mitigate it. This conclusion stems from a consideration of why implicit racial biases are concerning in the first place. To start, they might disproportionality increase the probability of a guilty verdict or a longer sentence for defendants of certain races. Second, they might play a role in judicial or juror decision-making despite being an extralegal factor. Accentuating race salience and social desirability effects, by leading white judges and jurors to consciously monitor their own behavior on the basis of potential racial bias, would likely ameliorate the first concern. However, the act of making race a central consideration would simultaneously intensify the second concern. Namely, by taking concerted efforts to not stereotype a black defendant, the race of the defendant inevitably becomes an explicit factor in a judge or juror's thought process.

Furthermore, by amplifying social desirability effects, some individuals might overcorrect for biases and redress neither the first nor the second concern. A case that proponents of raising awareness often use to bolster their argument actually provides an excellent example of this potential counter-effect. The study gave doctors a hypothetical scenario in which a patient exhibited specific symptoms, and subsequently asked the

doctors to rate the degree to which they would recommend a certain treatment (Green et al., 2007). For some, this hypothetical involved a black patient, while for others the patient was white. Doctors with stronger white preferences on the IAT, who (a) saw the version with the black patient and (b) were cognizant of the purpose of the experiment, were much more likely to support administering the treatment than their cohorts given the white patient. Importantly, these compensatory effects did not raise the level of recommendation to be equal to that given for white patients, but instead surpassed it. Thus, social desirability or race salience effects might just funnel the influence of bias in the opposite direction.

Even if raising awareness did promote less biased behavior, the impact of such an intervention relies heavily on active self-regulation and effortful cognitive strategies. Yet, as mentioned above, trials are inherently stressful and cognitively demanding situations, so the remaining resources available to vigilantly monitor one's biases would be quite limited. Without persistent self-regulation, significant reductions in implicit racial biases would be unlikely (Amodio, 2014; Banaji et al., 2003). Additionally, successfully monitoring one's biases might come at the price of not fully focusing on the facts of the case.

2.2.2 Screening Methods

Although screening methods would likely reduce the net level of implicit racial biases, there are multiple objections to be made. First, scholars caution against using the IAT as a diagnostic tool since it was designed to produce reliable results on an average group level, not to make predictions for specific individuals (Banaji et al, 2003; Kang et

al., 2012). Even so, incorporating screening methods might limit the pool from which potential judges or jurors can be drawn. To see how this might be the case, first consider that if such a policy were to be implemented, it would make sense to screen for those with moderate or strong preferences in order to substantially reduce the level of implicit racial bias in the courtroom, at least to a degree that would objectively make designing, introducing, and enforcing the policy worthwhile. However, data from an experiment surveying IAT scores across the United States suggests that the majority of people have more than just a slight bias, and that the average level of bias varies from state to state (Mooney, 2014; Xu, Nosek, & Greenwald, 2014). With judges, not only would screening narrow the pool of those eligible to hear a case, but this also might cause logistical issues since judges are already limited in number and have burdensome caseload pressures, particularly at the trial court level (Habel & Scott, 2014). In the case of jurors, if we think that citizens have a strong interest or moral right to serve on a jury (a question that will be addressed in Chapter Five), then we would be barring a large number of people from accessing this right, especially in a state with more severe biases on average. As the number of eligible jurors decreases, the ability of the jury pool to be representative of the community might also decline. Such a result would be problematic, as the jury pool is supposed to involve “a fair cross-section of the community” (see *Taylor v. Louisiana*, 1975 (and cases cited therein)). Although a pre-screened jury pool would be more impartial with respect to implicit racial biases, it might not be sufficiently representative or unbiased in other domains given its limited size. Additionally, requiring judges and jurors to take the IAT might raise privacy concerns, as individuals might not want the

courts to have information regarding their own degree of implicit racial bias (Roberts, 2012). Even though judges and jurors are already asked to divulge personal information during processes such as motions for recusal or voir dire, the fact that individuals are often not cognizant of their implicit biases might render IAT scores a further breach of privacy than is currently accepted. In other words, determining an individual's level of implicit racial bias goes past what the individual could reveal on his own, and exposes subconscious thoughts that are socially undesirable.

2.2.3 Weakening the Stereotypes

As noted earlier, proponents of using counter-stereotypical images to weaken negative associations often invoke Dasgupta and Greenwald's (2001) study involving famous historical figures. However, when a different group of researchers tried to replicate these results, they found that the stereotypes were harder to break than previously suggested (Joy-Gaba & Nosek, 2010). In order to figure out whether having examples of both positive black figures and negative white figures was necessary to achieve the effect, the researchers ran an experiment using solely counter-stereotypical exemplars of black individuals. Not only did the authors not observe any significant reduction in implicit racial biases following the intervention, but when they added disliked white figures back into the experiment, the reduction occurred with a substantially smaller effect size than Dasgupta and Greenwald (2001) had originally reported. Thus, using counter-stereotypical pictures to successfully reduce implicit racial biases against black individuals apparently requires encouraging negative stereotypes of white individuals (at least with this specific protocol). Implementing a strategy that

degrades one group in order to counteract the degradation of another group seems both counterintuitive and unwarranted, especially if the effects of this strategy are not as potent as presumed.

With respect to mental imagery exercises, although such tasks could theoretically constitute a viable solution, they are unlikely to be productive in the courtroom setting. For one, they rely extensively on the cooperative effort of individuals, and there might also be a large degree of variation in the content and vividness of these simulations from one person to the next (given idiosyncrasies inherent to imagination). To achieve a more standardized process, one might suggest simply showing judges or jurors a video. An experiment employing this approach found reduced IAT scores after participants watched video clips of black individuals in positive contexts, such as a family barbeque (Wittenbrink et al., 2001). However, only showing one race, even if it avoids presenting negative images of another race, positions racial factors as an obviously salient issue. Additionally, despite capturing the dynamic aspect of mental imagery in a systematized manner, the interactive effects of imaginative exercises, which are considered one of the most important features for enhancing learning (Vogel et al., 2006), are lacking in video-based strategies.

2.2.4 Diversification Efforts

The case for diversification in principle seems the most promising of the four proposed interventions, since the mere presence of someone with an additional perspective has been shown to have positive effects on group decision-making (DeGrassi, Morgan, Walker, Wang, & Sabat, 2012; Phillips, Northcraft, & Neale, 2006). However,

this type of intervention by itself would still be insufficient to adequately reduce implicit racial biases across trials. To start, diversification efforts would mainly be isolated to jurors, since most cases are heard by a single judge. Unless we overhaul the current system to make all trial courts involve a panel of judges (Rachlinski et al., 2009), making the pool of judges more diverse would have negligible effects, as individual judges would still lack the opportunity to have their opinions or decisions checked by a colleague of a different perspective during the trial. Put differently, increasing the diversity among judges might be beneficial for broadening viewpoints or attitudes in general, but cannot adequately address implicit racial biases when only one judge manages a case.

Despite the fact that group dynamics are a central feature of juries, several considerations complicate a proposal for diversification. Looking for diverse jurors requires a definition of diversity, and might lead to quotas or arbitrary decisions about who is diverse enough to facilitate the desired effects. Moreover, while heterogeneous groups tend to arrive at more carefully reasoned decisions than homogenous groups, having too many disparate or competing voices can hinder cooperation and thwart the original benefits of diversification (Maznevski, 1994). Furthermore, diversification interventions would be vulnerable to the issues discussed above regarding overcompensation due to race salience and social desirability effects. If white jurors start making concerted efforts to not appear racially biased to jurors of other races, they might interpret (or voice their interpretations) in a manner that goes beyond neutrality and instead constitutes bias in the other direction.

2.3 An Alternative Proposal: Neurointerventions

Implicit racial biases pose serious concerns for the justice system, yet the existing proposals for mitigation all appear to fall short. However, what if alternative methods could provide a more effective means to reduce implicit racial biases in both judges and jurors? Moreover, what if there was a way to avoid relying on hypervigilant self-monitoring, perpetuating race as an extralegal factor, or encouraging negative stereotypes?

While the science is still emerging, there is reason to believe that certain neurointerventions could meet these criteria in the future. Given the lack of available research involving neurointerventions in the courtroom setting, this paper will primarily refer to the use of neurointerventions as a theoretical proposition. However, in order to provide a more tangible conceptualization of this proposal, the following chapter will outline specific techniques that could constitute viable solutions pending further research.

Before delving into these potential mechanisms, a few disclaimers are in order. First, because research on the intersection between neuroscience and the mitigation of implicit racial biases is very much in its infancy, some of the methods discussed below have only been tested in a handful of studies with limited populations and constrained contexts. Second, while this paper suggests applying these mechanisms to the courtroom setting, none of the techniques have been explicitly designed, tested, or even envisioned for that particular purpose. Third, some proposals posited below are mere speculation of ways that certain technologies could be employed in the courtroom context, and have not been empirically tested (or at least not published in an available format). Consequently, while

the results and methodologies of existing studies will be explained, the following chapter does not intend to imply that any of the proposed neurointerventions are ready to be implemented, have been definitively proven effective, or would offer a foolproof solution. Instead, the goal of the subsequent section is to provide proof of concept, and present a carefully selected set of neurointerventions that has the potential to reduce implicit racial biases in the near future, without facing the same roadblocks as the proposals put forth in the existing literature.

3. Potential Neurointervention Techniques: A Proof of Concept

As discussed in previous sections, implicit racial biases stem from automatic and stereotyped responses to certain stimuli; although these biases can be mitigated through concerted efforts at self-regulation, the cognitively and emotionally demanding nature of trials hinders an individual's ability to engage in such monitoring. The neurointerventions proposed below aim to address both facets of this problem. On one hand, by targeting neural regions involved in the automatic processing of racial outgroups, neurointerventions might diminish the influence of learned stereotypes on subsequent judgments. On the other hand, implicit racial biases could also be reduced by focusing on areas implicated in monitoring and cognitive control, thereby enhancing self-regulation capabilities. This chapter will explain and examine two categories of potential neurointerventions that have the capacity to achieve these goals: computerized brain-training tasks and noninvasive brain stimulation techniques. An overview of the mechanisms will first be provided, followed by a discussion of safety and pragmatic factors. This chapter will then conclude by outlining the necessary steps that future research must take in order for neurointerventions to be implemented in the courtroom setting.

3.1 Overview of the Proposed Mechanisms

3.1.1 Computerized Brain-Training Tasks

While videos and mental imagery exercises would likely not be effective in the courtroom (for reasons explained earlier), computerized brain-training tasks could harness the benefits of these techniques in a manner appropriate for the intended setting.

To review, mental imagery exercises actively engage individuals in the simulation of counter-stereotypes in a way that is translatable to real-world situations. However, individual differences in imagination strategies might prevent these exercises from producing consistent results. Meanwhile, although videos can ensure that individuals all visualize the same material, they lack the interactive benefits that are crucial for learning. Computerized brain-training tasks have the potential to maintain the interactive and dynamic nature of mental imagery exercises while creating a standardized visual environment. These tasks could come in multiple forms, two of which will be postulated here.

First, judges or jurors could play a simulated game in which they must navigate a predetermined scenario or setting. Computer games could build on the design of the video-based study cited earlier (Wittenbrink et al., 2001) to more subtly present counter-stereotypical situations. For instance, an avatar could walk through a park, where families of *multiple* races are enjoying barbeques, reading on a bench, playing fetch with a dog, and so forth. The avatar could also be required to interact with various individuals, and these interactions could be structured to more frequently involve one's racial outgroup, albeit in a proportion that would not create race-salient issues. Additionally, the game could contain mandatory steps that the player must take in order to ensure engaged compliance with the task. Granted, this imagined game has not been studied empirically (or even developed), and would need to be thoroughly tested before further recommendations regarding its implementation could be made. However, given that it combines elements that are independently known to (a) reduce implicit racial bias, (b)

enhance learning, and (c) promote goal-directed behaviors, it seems reasonable to hypothesize that the game could be developed and provide successful results.¹

In addition to simulation games presenting counter-stereotypical scenarios, computerized tasks could also reduce the degree to which individuals view members of an outgroup as being significantly different than themselves. Given that implicit biases often stem from automatic reactions to potential threats, diminishing the “other-ness” factor of outgroup individuals might serve to weaken the implicit stereotypes (Terbeck et al., 2015). Researchers have employed this concept through body ownership illusion paradigms, in which participants are temporarily made to feel as though a body part belonging to someone of their racial outgroup is actually their own (Maister, Slater, Sanchez-Vives, & Tsakiris, 2015). After completing the task, not only do participants in these experiments consider the outgroup member to be more comparable to themselves, but they also demonstrate reduced IAT scores (Maister, Sebanz, Knoblich, & Tsakiris, 2013; Paladino, Mazzurega, Pavani, & Schubert, 2010).

In the courtroom context, these body ownership illusions could be presented in the form of virtual reality techniques. In immersive virtual environments, participants frequently use a head-mounted display that replaces all visual input from the actual surroundings with pre-designed simulations. Moreover, by tracking head and body movements, the virtual environment can allow participants to interact with the setting in a

¹ It is important to note that a similar gaming experience already exists as a civics education tool for young students. “iCivics,” a learning resource founded by former Supreme Court Justice Sandra Day O’Connor, involves online games pertaining to the justice system, including simulations of the juror experience (iCivics, n.d.). Future research could build on this initiative’s existing framework to facilitate the development of the game proposed in this paper.

way that mimics real life (Sanchez-Vives & Slater, 2005). In an experiment employing the self-other body distinction, Peck and colleagues (2013) had participants enter a virtual environment where a mirror reflected back an image of their avatar. When the participant moved his own limbs in front of the mirror, the reflection moved in an identically synchronous way. Light-skinned participants who were given a darker skinned avatar in the virtual setting demonstrated reduced implicit racial biases following the experiment (Peck, Seinfeld, Aglioti, & Slater, 2013). A similar technique could be especially potent in the courtroom setting, as it could combine the self-other body distinction with elements from the computer game discussed above. Judges and jurors could enter an immersive virtual environment and engage in the same interactions as those in the computer game (e.g., seeing people of various races in positive settings) via the perspective of an outgroup member. Importantly, the outgroup identity that the judge or juror takes on in the virtual environment would need to involve multiple characteristics so as to avoid race salience effects. For instance, the outgroup identity could entail someone of a different race, gender, and socioeconomic background (as portrayed through work uniforms (Freeman, Penner, Saperstein, Scheutz, & Ambady, 2011)), which could potentially make judges and jurors more open to outgroups in general. Computerized brain-training tasks could therefore dampen automatic reactions to members of other races in an interactive and efficacious manner.

3.1.2 Noninvasive Brain Stimulation

While computerized brain-training tasks function through actively engaging the individual in simulations, noninvasive brain stimulation techniques might offer a more

direct method of reducing implicit racial biases. Transcranial magnetic stimulation (TMS), a process in which a coil is placed on an individual's scalp, involves passing a current from a magnetic field through to specific brain regions near the scalp's surface (Farah, Smith, Ilieva, & Hamilton, 2014). Another procedure, transcranial direct current stimulation (tDCS), entails stimulating cortical regions via electrodes to a weaker degree than TMS (Farah et al., 2014). Researchers have found these methods capable of not just increasing or decreasing activation in particular areas and networks, but also modulating corresponding behaviors. In the courtroom setting, TMS and tDCS could enhance self-regulatory abilities or diminish stereotyped associations by either targeting neural regions such as the medial prefrontal cortex (mPFC), the anterior temporal lobe (ATL), and the dorsolateral prefrontal cortex (dlPFC), or by focusing on the larger networks involved in these processes. Let us start by examining the first route.

The mPFC, which has been implicated in cognitive control, social perception, humanization, and perspective-taking, is known to be less active during biased judgments (Amodio, 2014). Accordingly, a recent study employing tDCS found that increasing activity in the mPFC reduced participants' scores on the IAT compared to participants with decreased or unaltered mPFC activity (Sellaro et al., 2015). Next, the ATL is believed to play a critical role in stereotyped mental representations (Amodio, 2014), and researchers have found that applying inhibitory TMS to either the left or right ATL significantly decreases participants' IAT scores (Gallate, Wong, Ellwood, Chi, & Snyder, 2011). Lastly, the dlPFC is involved in self-regulation, cognitive control, and goal-directed behaviors (Amodio, 2014). While studies stimulating the dlPFC to modulate

racial biases appear to be lacking, researchers have explored the effects of TMS on gender stereotypes, and found that applying TMS to the left dlPFC and the right anterior dlPFC significantly reduced implicit gender bias scores (Cattaneo, Mattavelli, Platania, & Papagno, 2011). Future studies would be necessary to test whether TMS in the dlPFC would modulate implicit racial biases as well. Moreover, as research expands in this domain, stimulating other brain areas implicated in cognitive control and stereotyped judgments could also hold promise. These areas might include the inferior frontal gyrus, which is involved in response inhibition, the temporoparietal junction, which deals with self-other perceptions, and the fusiform gyrus, which plays a role in processing outgroup faces (Amodio, 2014; Banissy & Ward, 2013; Terbeck et al., 2015). It is important to note, however, that the brain regions associated with forming or regulating implicit racial biases are also implicated in a wide array of other cognitive processes. Thus, while this area of research has produced promising preliminary results, the nature and extent of collateral effects from general stimulation to these regions would need to be examined.

Although studies exploring distributed networks have not been conducted with respect to implicit racial biases in particular, focusing stimulation efforts on connected neural regions might offer more precise results. Interestingly, researchers have discovered brain-state-dependent effects of noninvasive brain stimulation, finding that activating networks involved in a task prior to or during stimulation can increase the potency and selectivity of the intervention's effects (Jacquet & Avenanti, 2013; Sergeeva, Henrich-Noack, Bola, & Sabel, 2014). Thus, future studies might ask individuals to envision positive counter-stereotypes or engage in self-regulatory mental imagery exercises during

stimulation so as to target the population of neurons actually involved in such cognitive processes. Additionally, researchers have found that if applied to the correct site, noninvasive brain stimulation can modulate connectivity within distributed networks to enhance or inhibit performance (Fox, Halko, Eldaief, & Pascual-Leone, 2012). Since research has elucidated networks involved in both the regulation of outgroup biases as well as the formation of stereotypes themselves (Amodio, 2014), targeting these pathways (through excitatory stimulation to the former and inhibitory stimulation to the latter) might produce results that are more specific to implicit racial biases.

Research on noninvasive brain stimulation methods, specifically in the context of implicit racial biases, is still an emerging discipline. Not only is our understanding of the topic continually evolving, but extensive research would also be required to validate the hypotheses put forth in this chapter. However, given preliminary findings as well as the pace of development in the field, noninvasive brain stimulation techniques have the potential to reduce activation in areas involved in stereotypic associations while increasing activity in regions necessary for self-monitoring and enhanced vigilance.

3.2 Relevant Safety and Practical Considerations

The goal behind implementing neurointerventions in the courtroom is to reduce implicit racial biases better than the previously suggested methods. Therefore, any intervention that put individuals at risk or did not hold potential to effectively or feasibly make a difference would be undesirable. Accordingly, it is important to identify the relevant safety and pragmatic factors inherent to any proposed neurointervention. The following section will address these considerations for computerized brain-training tasks

and noninvasive brain stimulation techniques.

3.2.1 Computerized Brain-Training Tasks

The safety risks associated with computerized brain-training tasks depend on the proposed exercise in question. The risk profile for computer simulation games is likely identical to that of daily computer usage, including mild symptoms such as eyestrain and muscle fatigue (Hayes, Sheedy, Stelmack, & Heaney, 2007). Immersive virtual environments could present an added psychological side effect of unease due to the changes in perceived body ownership. This method might also produce slight motion sickness or discomfort from using the head-mounted display (Lawson, 2014).

With respect to pragmatic considerations, computerized brain-training tasks would likely take a very brief time to complete, perhaps even in the range of 20 minutes or less. For instance, even though immersive virtual environments require extra time to set up the head-mounted display and orient the individual, studies involving this technique often last for only a few minutes once the participant enters the virtual environment (Ahn, Bailenson, & Park, 2014; Peck et al., 2013). One important area for future research with immersive virtual environments will be to determine the most simple and cost-effective means capable of achieving the desired results. For example, instead of placing individuals in a room with four projection walls and objects that they can reach out to touch (which is commonly the case in virtual reality paradigms), a basic setup involving a head-mounted display, headphones for sound perception, and a joystick to navigate the virtual setting might be adequate (Sanchez-Vives & Slater, 2005). Should such a setup be sufficient, virtual reality techniques would be a cost-effective option; one device could be

used by multiple individuals, and head-mounted displays are now being sold as consumer products for approximately \$250, a price that will likely decrease as the devices become more widespread among the broader consumer community (Poon, 2015).

While the implementation of computerized brain-training tasks will depend on insights from future research efforts, these techniques could be applied to the courtroom setting in the following ways. First, since computer simulation games would involve limited time constraints, judges or jurors could perform these tasks before each day of the trial or between stages such as jury selection and opening statements (should there be enough time between them). Second, given that immersive virtual environments could generate longer-lasting results due to their more vicarious and interactive nature, it might suffice to employ this method in the form of training programs. In the case of judges, virtual reality environments could be introduced into continuing judicial education seminars. These tasks could be made especially intricate and involved so as to produce effects potent enough to last for a period of months. A modified version of this virtual environment could be created for jurors, who could engage in the task immediately after being selected for service. This variation of the task could be designed to induce more temporary effects on a timeline consistent with that of a typical trial, given that jurors are only on duty for a limited period of time.

3.2.2 Noninvasive Brain Stimulation

Turning to noninvasive brain stimulation, both TMS and tDCS involve relatively minimal risks (Farah et al., 2014). Possible side effects for TMS and tDCS include discomfort, a tingling feeling, skin irritation, and lightheadedness; early uses of TMS

reported the risk of seizures, but as the technology has advanced, this risk is considered quite rare (Brunoni et al., 2012; Farah et al., 2014; Mayo Clinic Staff, 2015). However, the technology is unadvised for individuals with certain characteristics, such as having metallic neural implants, and the long-term effects of both methods are unknown (Brunoni et al., 2012). While one might worry that the potential side effect of headaches or lightheadedness could detract from a judge or juror's ability to pay attention during the trial, these side effects are generally very temporary, and would likely subside in the interim between stimulation and the commencement of the trial.

Regarding practical considerations, noninvasive brain stimulation techniques would take approximately half an hour to conduct. In the TMS and tDCS experiments examining implicit racial biases, participants underwent stimulation for about 15-20 minutes (Gallate et al., 2011; Sellaro et al., 2015), not including the time necessary to set up the procedures (~10 minutes). Similar to virtual reality techniques, the equipment for these technologies, especially tDCS, is becoming more mainstream and affordable (Jwa, 2015), and can be reused many times. Additionally, while studies have not explicitly looked at the duration of effects over a time period similar to that of a typical trial, it appears that effects are most pronounced directly after stimulation (Brunoni et al., 2012); moreover, whereas single sessions can in some cases produce effects lasting for 24 hours (Falcone, Coffman, Clark, & Parasuraman, 2012), if long-lasting results are desired, repeated sessions over the course of days or weeks are advised (Nitsche et al., 2008).

With respect to potential implementation strategies, noninvasive brain stimulation could be administered prior to segments of the trial where increased vigilance might be

particularly important, such as before opening statements (when initial impressions are formed, and when stereotypes are likely to play a crucial role), or before conviction and sentencing decisions (when the judge or jury's decision-making process directly impacts the outcome of the trial). In the case of judges, who often attend multi-day continuing education seminars, noninvasive brain stimulation could be incorporated into judicial training programs to achieve longer-lasting effects.

3.3 The Need for Tailored Research Efforts

As previously stated, research on the intersection between neurointerventions and implicit racial biases is still in nascent stages, and has not been explored in a setting even remotely similar to that of a courtroom. Although computerized brain-training tasks and noninvasive brain stimulation techniques hold potential to be effective tools in the future, tailored research efforts are necessary to move these neurointerventions from theoretical possibilities into viable solutions. Specifically, future experiments will need to address (1) the strength, scope, and duration of the neurointerventions' effects, and (2) the degree to which individual differences produce variable results.

To start, while studies have begun to examine the impact of these two neurointerventions on reducing implicit racial bias scores, researchers will need to elucidate the extent of bias reduction on factors beyond just performance on the IAT. In other words, computerized brain-training tasks and noninvasive brain stimulation would have to significantly reduce biased assessments of evidence or automatic presumptions of guilt. In addition to potentially decreasing implicit racial biases, neurointerventions would need to achieve these results without creating collateral impacts, such as altering other

cognitive processes or modulating explicit beliefs. Otherwise, unintended consequences could nullify the neurointerventions' benefits by producing counterproductive effects, which in turn might jeopardize the trial process. Within the analysis of strength and scope is also the question of the effects' durations. Future research would need to determine the longevity of the effects, and acceptable standards for durability might differ between judges and jurors. For instance, judges are consistently in the courtroom setting and would have a continuous need for bias reduction. In contrast, jurors are only responsible for meeting their required duties of impartiality for a limited amount of time, and thus the neurointerventions' effects should be keyed to the length of the trial. Moreover, future research would need to identify which variations of the techniques would be the most efficacious, simple and cost-sensitive to implement. For example, as mentioned above, if a basic setup for immersive virtual environments could meet the desired strength, scope, and duration requirements, then such a method would be preferable compared to more complex or resource-constraining options.

Not only are questions of strength, scope, and duration essential to resolve, but in order for neurointerventions to be useful, they would also have to reliably produce the intended effects in the majority of individuals. If the techniques were only successful in a small portion of the population, the overall reduction in implicit racial bias might be too subtle to make the costs of implementation worthwhile. Thus, systematic research would need to specifically examine the extent to which individual differences might modulate results.

One place to begin this research could be to replicate the studies mentioned in Chapter One using mock trial scenarios, and explore whether and to what degree training with the proposed mechanisms reduces implicit racial biases (compared to a control group). Such a study could be expanded to involve multiple versions of the techniques ranging in levels of complexity or resource-constraints, as well as include individuals of different ages, races, and backgrounds. These studies would also need to entail follow-up tests to provide longitudinal data on the neurointervention's effects.

3.4 Neurointerventions as a Feasible Prospect

Given the hardwired nature of stereotypes, as well as the cognitively demanding aspects of the trial environment, it is clear that concerted strategies will be necessary to overcome implicit racial biases in the courtroom. Although research on computerized brain-training tasks and noninvasive brain stimulation techniques is currently limited, both hold the potential to either diminish the potency of racial outgroup stereotypes or enhance self-regulatory abilities. Whereas existing suggestions for reducing implicit racial bias in the courtroom are either unlikely to produce significant results, make race too salient an extralegal factor, or encourage negative counter-stereotypes, these neurointerventions could avoid such pitfalls.

The purpose of this chapter was to (1) outline two neurointervention methods that could serve as valuable and feasible mitigation strategies in the future, and (2) to demonstrate that the prospect of neurointerventions to reduce implicit racial biases has a scientific basis grounded in preliminary experimental findings. Although this chapter has argued that both computerized brain-training tasks and noninvasive brain stimulation

techniques could be applied to the courtroom setting, computerized brain-training tasks are likely the more plausible and promising option. Since computer games and virtual reality techniques are more thoroughly researched, mainstream, and widely used, their practical application might be more obtainable given society's increased familiarity with the general methods. Nevertheless, both techniques serve to provide a proof of concept.

In sum, computerized brain-training tasks and noninvasive brain stimulation could become viable strategies in the future to reduce implicit racial biases in the courtroom; however, due to the degree of research yet to be conducted, the rest of this paper will examine neurointerventions as a general and theoretical technique. For the remainder of this discussion, let us assume that in the near future, available neurointerventions will be safe, effective, limited in scope, and feasible to implement in the courtroom setting.

4. Neurointerventions in Context: The Case of Judges

Now that the problem of implicit racial biases in the courtroom has been explained, and the use of neurointerventions to alleviate the issue has been proposed, it is pertinent to examine how such methods could be applied to judges presiding over criminal cases.¹

Judges hold a unique position in the legal realm, as they not only choose to enter into their professional role, but also serve as the tangible face of the justice system.

Additionally, judges play a prominent part in influencing trial outcomes through managing the trial, determining the admissibility of evidence and appropriateness of certain statements, and ultimately making the sentencing decision in most cases. Given judges' direct involvement in each case they hear, problems stemming from biased judgments are particularly concerning. Moreover, the strength of stereotypical associations between criminality and race might be even more potent in judges than in the general public; since black individuals comprise a disproportionate percentage of the defendant population, judges might possess skewed perceptions of criminal trends (Papillon, 2013).

The following chapter will examine the prospect of using neurointerventions to reduce implicit racial biases in judges from a variety of angles. First, this chapter will explain what a system with neurointerventions would look like, including proposals for both a voluntary and mandatory policy. Next, ethical and legal considerations will be explored for the two policies, and arguments for the permissibility of both systems will

¹ Although implicit racial biases are also a problem in civil cases, the present discussion will focus solely on criminal trials.

be put forth. This chapter will then go one step further to contend that judges have a moral obligation to participate in the neurointerventions.

4.1 Neurointervention Policies: Painting the Picture

Implementing neurointerventions in the courtroom setting could potentially take one of two forms. Let us first consider a voluntary policy. In this system, judges would be made aware of the problem of implicit biases in the courtroom and be given information about the neurointerventions, including facts about their safety and efficacy.² Judges would then have the option of partaking in the neurointerventions prior to trials. In contrast to a voluntary policy, a mandatory process would make neurointerventions a requisite aspect of the judicial profession. Since individuals are not required to become judges, this policy will be referred to as “pseudo-compulsory” from hereon out; individuals would not be forced to participate in the neurointerventions, but would have to agree to them in order to enter or remain in the profession. The subsequent sections will now address the ethical and legal issues associated with voluntary and pseudo-compulsory neurointerventions.

4.2 A Principle-Based Approach

Given that individuals would be engaging in procedures primarily intended to serve the broader society, it is prudent to assess the associated benefits and burdens through the

² In order to avoid any race salience effects that might arise in explaining the reason for neurointerventions, judges would likely be informed of the problem of implicit bias as a general phenomenon. This instruction should not be viewed as deceptive, since even though the interventions are targeted at implicit racial biases in particular, the enhancements in self-regulatory capacities and reductions in automatic self-other distinctions are generalizable and impact the fundamental processes underlying many types of implicit biases.

lens of a principle-based bioethical framework. The following section will provide an overview of four foundational principles: autonomy, nonmaleficence, beneficence, and justice (Beauchamp & Childress, 2013). This section does not purport to offer a comprehensive or in-depth explanation of these principles, but instead delineate the central issues at stake. Accordingly, the basic tenets of each principle will first be outlined, and the application of these concepts to the context of neurointerventions will be addressed in the respective sections on voluntary and pseudo-compulsory policies (Sections 4.3 and 4.4).

4.2.1 Autonomy

While the meaning of autonomy can take multiple forms in bioethical discourse, there are two senses of the word that are particularly pertinent for the current discussion. First, the term autonomy is frequently used in relation to an individual's decisions, and expresses the idea that people should not only have a voice in matters concerning their own wellbeing, but also that their decisions should be respected (Feinberg, 1986). Since thoughts and opinions inherently guide an individual's decisions, this concept of autonomy also encapsulates the preservation of freedom of thought, which maintains that people have a moral right to think or believe whatever they desire, regardless of whether these thoughts are fervently held opinions or subconscious preferences (United Nations General Assembly, 1948). Backing up one step further, respecting a person's autonomous decisions and thoughts also requires an absence of interference with the mental processes that give rise to those beliefs (Bublitz, 2015; Bublitz and Merkel, 2012).

The second prong of autonomy involves the decision-making process itself. Autonomous decision-making requires that individuals be free to form decisions on their own accord without undue external influence (Feinberg, 1986). When decisions are made under pressured or manipulative conditions, concerns of coercion often arise. Specifically, coercion entails three main factors: (1) the reduction of an individual's available choices, (2) a detrimental effect on this individual as a result of the constrained choice set, and (3) manipulative intent on behalf of the agent responsible for the restrictive situation (Hawkins & Emanuel, 2005). In coercive circumstances, an individual's autonomous decision is said to be undermined or invalidated, since the decision might not be fully reflective of his own will (Beauchamp & Childress, 2013).

4.2.2 Nonmaleficence

Even if a person autonomously provides his consent, the principle of nonmaleficence asserts that interventions or policies should avoid creating harm, especially to the individual participating in the procedure. Importantly, this potential harm is not just confined to negative impacts on a person's health, but also includes harm to one's interests, mental state, or general functioning. While unexpected outcomes are always a possibility, nonmaleficence entails identifying and minimizing foreseeable risks (Beauchamp & Childress, 2013; Levine, 1988).

4.2.3 Beneficence

Beneficence requires that in addition to preventing harm, the intervention or policy in question should also confer a considerable degree of benefit. This benefit can be in relation to the specific individual directly impacted by the intervention, or refer to

advantages brought to the broader society. Moreover, the principle of beneficence involves a risk-benefits analysis, in which the overall benefits must outweigh any harms induced (Beauchamp & Childress, 2013; Levine, 1988; Presidential Commission for the Study of Bioethical Issues, 2015).

4.2.4 Justice

Lastly, the justice principle typically entails both comparative and distributive components (Levine, 1988). First, comparative justice necessitates that individuals in similar situations be treated in a similar manner. Second, while there are various theoretical interpretations as to how this distribution should occur, the basic premise of distributive justice requires that certain groups within the community do not disproportionately reap the benefits or bear the burdens produced by the intervention (Levine, 1988; Presidential Commission for the Study of Bioethical Issues, 2015).

Now that the underlying considerations involved in each of the four principles have been enumerated, let us look specifically at the voluntary and pseudo-compulsory policies to evaluate whether these principles are upheld or compromised.

4.3 The Ethics of Voluntary Neurointerventions

4.3.1 Autonomy

To review, the voluntary policy would provide judges with the option of participating in the neurointerventions, without imposing any consequences for declining to do so. Since judges would be given this unconstrained choice, the policy does not appear to undermine a judge's ability to make an autonomous decision to either consent to the neurointerventions or forego the opportunity.

While a judge's ultimate decision might be unequivocally respected in a voluntary policy, the more pressing issue is freedom of thought. It is important to note that if the neurointerventions function as expected, they will not implant thoughts into individuals' minds, erase all opinions about race, or dictate what people should believe. Accordingly, they should not be viewed as *fundamentally* altering one's personality, attitudes, or opinions, especially since explicit beliefs can often be inconsistent with one's implicit biases. Instead, the neurointerventions would modulate processes that influence the nature or likelihood of certain automatic judgments. In fact, multiple scholars have contended that such automatic processes, like implicit biases or impulses, actually decrease one's capacity for mental control and therefore impair freedom of thought and autonomy in their own right (Bublitz, 2015; Douglas, 2013). In other words, the presence of implicit biases, particularly when they are divergent from one's explicit attitudes, can cloud a person's ability to thoroughly exercise his consciously-held beliefs or opinions. Nevertheless, implicit biases can still be considered thoughts (or a component of one's thoughts), and by altering the neural processes leading to such biases, neurointerventions would constitute a threat to freedom of thought.

Despite the policy's interference with freedom of thought, it is important to remember that autonomy also entails respect for an individual's decisions. Although the neurointerventions would modulate the neural processes involved in the formation of implicit racial biases, they would not be impeding the ability of an individual to decide whether or not to participate in the policy. Thus, if judges voluntarily waive their freedom of thought in order to engage in the neurointerventions, and they do not make this choice

under coerced circumstances, then a policy honoring this decision would still be in accordance with the principle of autonomy.

4.3.2 Nonmaleficence

Given that the current discussion involves a judge's voluntary and optional decision to participate in the neurointerventions, the creation of the policy itself does not induce any apparent harm to the judge, since he would not face any consequences for declining to provide his consent.

For judges who choose to participate in the neurointerventions, the issue of nonmaleficence would largely depend on the specific mechanisms that are developed. However, as stated in the previous section on tailored research efforts, in order for any neurointervention to be implemented in the courtroom setting, the method would need to entail minimal safety risks and be limited in scope. The threshold for these requirements would likely be determined by a combination of technological capabilities and prevailing community sentiments. For example, let us imagine that the safety risks of immersive virtual environments were confined to slight muscle or vision fatigue, and that the technology not only reduced the strength of preexisting outgroup stereotypes, but also enhanced one's positive feelings towards outgroups in general. Those making decisions governing the judicial profession (such as the American Bar Association) would need to decide whether this degree of risk and scope would be acceptable. The larger judicial community would likely play a role in this decision as well given that enough of them would have to be amenable to the idea in order to make instituting the policy worthwhile. Since the neurointerventions would only be implemented if they met agreed-upon

standards, the voluntary policy does not appear to pose significant threats to the principle of nonmaleficence.

4.3.3 Beneficence

Calculations of beneficence include benefits granted to both the individual and the broader society. Let us analyze each in turn. With respect to the specific judge partaking in the neurointerventions, such an action would help him to better fulfill his obligation to impartially hear the case. Not only is this a professional duty (Special Committee on Standards of Judicial Conduct of the American Bar Association, 2014), but the inability to be impartial also constitutes legal grounds for disqualification from a trial (28 U.S.C. §455). Therefore, participating in the neurointerventions could conceivably allow the judge to more adequately meet moral, professional, and legal duties. To the extent that individuals derive benefits from better performance in their chosen line of work, the judge might gain personal satisfaction from more adeptly carrying out his role.

Turning to broader notions of beneficence, if judges choose to participate in the neurointerventions, both society at large and its individual members would benefit as well. Since trials are supposed to be impartial, attenuating biases and increasing fairness would help the system to serve its intended purpose. The trial system would therefore achieve greater legitimacy, which could bolster society's support for the institution. With respect to defendants, facilitating fairer trials through reducing implicit racial biases might make a defendant more likely to *actually* be presumed innocent from the beginning of the hearing. Moreover, since any individual could one day be a defendant,

neurointerventions could provide future benefits to all citizens should they find themselves in this position.

4.3.4 Justice

Within the context of the voluntary policy, the justice principle can be analyzed with respect to both the judges participating in the neurointerventions as well as the overall effects of the policy.

Since the policy would provide all judges with the same opportunity to participate, concerns with dissimilar treatment among similarly situated individuals would not arise. Additionally, the extent to which the benefits and burdens of the policy are distributed might depend on how many judges decided to consent to the neurointerventions; however, at the very least, the policy would not disproportionately impact certain individuals from the outset given that there are no set criteria for which judges can or cannot engage in the neurointerventions.

In terms of the policy's effects, the neurointerventions would facilitate comparative justice better than the status quo, while also meeting distributive justice standards. Starting with comparative justice, if two cases differed only with respect to the defendant's race, then a judge's choice to partake in the neurointerventions would enhance the likelihood that these defendants would be treated similarly, especially in regard to their access to a fair and impartial trial. Turning to distributive justice, while only those defendants whose cases were heard by a less biased judge would enjoy the benefits brought by the policy, the policy itself would not be systematically providing advantages or disadvantages to any particular group within the defendant population. The

reason for this is that since all judges would have the same opportunity to participate, and provided that judges are generally assigned to cases in a random manner, defendants would theoretically have an equal chance of having their case heard by a judge who consented to the neurointerventions.

4.3.5 The Ethical Permissibility of Voluntary Neurointerventions

As evidenced by the analysis above, a voluntary policy that offered judges the choice of participating in the neurointerventions would uphold each of the enumerated bioethical tenets; since the voluntary policy would respect a judge's autonomy, avoid creating undue harm, provide benefits to individual judges and the broader society, and abide by comparative and distributive notions of justice, the policy should be regarded as ethically permissible.

4.4 The Ethics of Pseudo-Compulsory Neurointerventions

If we agree that a voluntary neurointervention policy is ethically justifiable, let us now take the next step and consider the ethical permissibility of a pseudo-compulsory scheme (in which neurointerventions were a requisite aspect of the judicial profession). Whereas some of the considerations are identical to those addressed in the voluntary case, the mandatory nature of the pseudo-compulsory policy raises additional concerns. The following section will evaluate the relevant issues for each principle.

4.4.1 Autonomy

Given that the pseudo-compulsory and voluntary policies would modulate mental processes in the same manner, the discussion of freedom of thought is largely equivalent for both situations. However, the more pertinent question with the pseudo-compulsory

scheme is whether the policy promotes autonomous decision-making. While the proposed policy would not force individuals to engage in the neurointerventions, it would require judges to make a consequential decision. The argument might be made that such a policy would lead to coerced instances of consent, as a judge with a strong interest in entering or remaining in the profession might have no choice but to overlook his preferred desire not to participate in the neurointerventions. Hence, despite technically agreeing to the neurointerventions, this act might be suspect on the grounds that it was not based on a fully autonomous decision.

As previously articulated, coercion usually involves restricting an individual's choices in such a way that leaves the person less well-off (Hawkins & Emanuel, 2005). Moreover, scholars often distinguish between offers (which add options) and threats (which take them away), and claim that only threats by definition can constitute coercion; their reasoning is that offers would make the person better off by enhancing his choice set, regardless of whether the added options were desirable (Hawkins & Emanuel, 2005; Wertheimer, 1987). Additionally, threats entail an element of manipulation, in which the entire point of limiting the available choice set is to get someone to behave in a particular way (Hawkins & Emanuel, 2005). Given this definition of coercion, let us now apply it to the policy in question to see if and how the necessary criteria could be met.

In regard to the first requirement, that of reducing available options, one could claim that the judge's choice set has indeed been limited. Whereas current or aspiring judges previously had the option to serve as a judge without agreeing to the neurointerventions, this choice is no longer available. However, it is not entirely clear that a judge's choice

set has been narrowed as opposed to merely complicated through additional conditions. It is important to consider the fact that professions contain a set of requirements with which its members must comply, and failure to meet such obligations bars a person from entering or remaining in that occupation. Consequently, the choice set of judges can be viewed as (a) comply with the obligations required for the profession or (b) lose your qualification to hold this position. Although the stipulations have changed under the pseudo-compulsory policy, this overall option set stays the same.

Whether or not we believe that options have been reduced, the more questionable claim is that judges are worse off as a result of the policy. One way to address this issue is to apply Wertheimer's (1987) framework for considering whether a given choice set is coercive. Wertheimer claims that if a person's welfare has been decreased, then his "moral baseline," as defined by the moral rights to which he is entitled, has been lowered and a threat has been made. As aforementioned, even though offers might be morally problematic for other reasons, many scholars believe that only threats and not offers can be considered coercive.³ Hence, if the one creating the choice set lowers the moral baseline of the person choosing, then the act could be viewed as coercive; if the choice makes someone better off and thus raises his moral baseline, then the action is actually an offer and cannot be coercive. With the pseudo-compulsory policy, the created choice set is essentially "participate in the neurointerventions, or forfeit the opportunity to serve as a judge." Accordingly, assessing moral baselines will depend on whether there is a right to

³ It is acknowledged that not all scholars ascribe to the distinction between offers and threats. However, given that this categorization appears to be predominant in the literature on coercion, this framework will be employed for the purposes of this discussion.

be a judge.

While negative consequences might occur for a judge by deciding not to engage in the neurointerventions (an issue that will be discussed in the section on nonmaleficence), it seems difficult to say that a legal or moral right has been violated. To start, there is no legal right to be a judge, let alone to have any particular job. Moreover, as stated above, every profession contains a set of rules that its members are expected to follow, simply by virtue of being in that profession. Judges choose to accept these obligations, and would only be faced with this pseudo-compulsory decision if they voluntarily elected to join the profession, which is something they are in no way compelled to do, nor are entitled to do. Returning to Wertheimer's (1987) framework, if judges do not have a right to serve (which this paper suggests they do not), then their moral baseline has not been decreased, and a threat has not been made.

With respect to the third requirement of manipulation, one could theoretically contend that the government creates this restrictive choice set for the sole purpose of facilitating its desired outcome. If we assume that judges have a strong interest in keeping their position, then individuals might consent to the neurointerventions (even against their preferred wishes) in order to qualify for the profession. In other words, the threat of taking away the option to be a judge unless an individual agrees to the neurointerventions could be regarded as a manipulative act—more judges might consent to the neurointerventions to avoid forfeiting their eligibility, and as a result directly serve the interests of the government. However, such claims are not entirely convincing. While judges might weigh the use of neurointerventions differently than they would if the

consequence did not entail losing their job, the purpose of the policy and the intentions behind it are not focused on making judges succumb to the government's desires. Instead, the policy is designed to enhance the fairness of the justice system, create positive benefits for citizens, and enable judges to more adequately meet their professional duties.

In sum, while one could make the claim that the pseudo-compulsory policy is coercive, this section has argued that there are multiple reasons to doubt such a stance, at least according to the generally accepted definition of coercion employed in this discussion. As a result, if judges are not coerced into agreeing to the neurointerventions, then the pseudo-compulsory policy is in accordance with the respect for autonomy.

4.4.2 Nonmaleficence

The assumptions regarding the neurointerventions' safety and scope are the same as those put forth in the voluntary case. However, whereas the voluntary policy did not impose any consequences on a judge for declining to consent to the neurointerventions, the pseudo-compulsory policy would prohibit him from entering or remaining in the profession. Although the previous section argued that there was no right to be a judge, it is important to recognize that negative consequences might nevertheless ensue. For example, in the case of judges currently sitting on the bench, one could say that making them resign from their position for failure to consent to the neurointerventions would have a negative impact on their lives; they would lose their job, their financial security, their self-identity, and the ability to work in a role about which they might be fervently passionate. For those aspiring to enter the profession, not providing their consent and forfeiting their eligibility might undermine all of the effort and resources that these

individuals had invested in order to one day have the opportunity to serve as a judge. Thus, even though the policy would be minimizing the risk of physical harm by only offering neurointerventions that met acceptable thresholds of safety and scope, potential harm to a judge's broader interests might also exist.

4.4.3 Beneficence

The benefits produced by a pseudo-compulsory policy would be more or less equivalent to those enumerated in the discussion of the voluntary case: the judge himself would be better able to fulfill his professional duties, defendants would receive a more impartial trial, and the justice system would gain legitimacy through better meeting its promise of fairness. In fact, these latter benefits might be even greater with the pseudo-compulsory policy, since implicit racial biases would be reduced in all trials as opposed to just those in which the judge elected to partake in the neurointerventions.

4.4.4 Justice

As with the analysis for the voluntary policy, it is important to evaluate comparative and distributive justice considerations for both the judges themselves as well as the policy's resulting effects.

In the case of judges, requiring all members of the profession to take the neurointerventions would align with comparative justice standards, since every judge would be treated in the same manner with respect to the policy. Similarly, if each judge participated in the neurointerventions, the burdens associated with the tasks would be shared by all in the profession. With the voluntary policy, this paper asserted that although certain judges would not be disproportionately impacted from the outset, the

degree to which the benefits and burdens would be distributed would depend on the number of judges who decided to provide their consent. To see the advantages of a pseudo-compulsory policy in regard to distributive justice concerns, imagine that in the voluntary system, only a few judges chose to participate in the neurointerventions. If we agree that less biased judges would yield advantages to the justice system and enhance the legitimacy of the judicial profession, only those who engaged in the neurointerventions would be supplying these benefits. While these judges would be taking on an extra burden, those who refused to participate in the neurointerventions would reap the benefits of this fairer trial system without providing the necessary contributions themselves. Thus, one might claim that these latter judges would be unjustly free-riding on the actions of those who did partake in the neurointerventions (Wertheimer, 2002). With a pseudo-compulsory policy, the integration of neurointerventions into professional requirements would alleviate such concerns, as each judge would have a role in bringing about the policy's benefits.

While a pseudo-compulsory policy might distribute benefits and burdens among judges more than a voluntary policy, a pseudo-compulsory framework would also enhance comparative and distributive justice with respect to defendants. To start, every case that came through the criminal justice system would be heard by a judge with reduced implicit racial biases. As a result, each trial would have a greater chance of meeting standards of fairness and impartiality than it otherwise might have. Thus, not only would every defendant be treated similarly with respect to the policy, but the benefits would also not be confined to those defendants who happened to be assigned a

judge who consented to the neurointerventions.

4.4.5 The Ethical Permissibility of Pseudo-Compulsory Neurointerventions

While a principle-based approach to bioethics necessitates respect for the four central tenets, it is not uncommon for principles to conflict with one another. In the present discussion, the pseudo-compulsory policy would generate substantial benefits for the defendant, the trial system, and society at large. It would also meet demands of both comparative and distributive justice. However, if a judge declined to provide his consent, the policy might cause harm to his personal interests by taking away his ability to serve in that role.

Although these negative consequences should be taken seriously, it is important to note the particular nature of the judicial profession. By deciding to become a judge, individuals choose to abide by obligations such as exhibiting impartiality and facilitating a fair trial. At its core, the pseudo-compulsory policy would be promoting these very goals and standards. Therefore, it seems counterintuitive that the main issue with nonmaleficence concerns the infringement upon an individual's interest in being a judge, and yet the policy would be enabling judges to better satisfy the demands that encompass the fundamental aspects of the role. Moreover, judgeship is a professional occupation, and individuals are generally not entitled to retain their position if they cannot meet the standards that the profession requires. In other words, while the negative consequences that might result from declining to provide consent should not be overlooked, the volitional and professional nature of judgeship should not be discounted either. Additionally, the principle of beneficence involves a risks-benefits analysis and requires

that the benefits outweigh the risks. This section has argued (1) that the risk of harm to the judge must be considered within the context of his volitional choice to serve in this profession and abide by the associated requirements, and (2) that substantial benefits would be brought to defendants and the trial system as a whole (in addition to enabling the judge to better fulfill his own duties). As such, this paper contends that a pseudo-compulsory policy would be justifiable due its respect for autonomy, the significant advantages that it generates, and the degree to which it honors comparative and distributive justice standards.

4.5 Neurointerventions as a Moral Obligation

Now that a case has been made for the ethical permissibility of voluntary and pseudo-compulsory policies, let us turn to the main contention of this paper: namely, that if we assume neurointerventions will be effective, feasible, and offer better solutions than existing mitigation strategies in the future, judges have a moral *obligation* to participate in these neurointerventions to reduce their own implicit racial biases.

This moral obligation stems from both professional responsibilities and duties to others. The requirement to impartially hear cases is an expectation made clear in both judicial codes of conduct as well as state and federal laws (e.g., 28 U.S.C. §455; N.C. Gen. Stat. §15A-1223; Special Committee on Standards of Judicial Conduct of the American Bar Association, 2014). Given that (1) implicit racial biases are known to be prevalent among judges (Rachlinski et al., 2009), and (2) these biases have demonstrably negative impacts on impartiality, failure to alleviate the problem would compromise a judge's ability to fulfill his duties. As a result, if judges have a professional and legal duty

to reduce mutable impediments to their impartiality, and if neurointerventions are the only or most effective way to achieve this goal, then it follows that judges have a moral obligation to participate in the neurointerventions. Importantly, this duty is not just grounded in professional obligations, but is also rooted in the legal rights of a defendant. The Fifth and Sixth Amendments to the Constitution entitle defendants to fair criminal justice proceedings with due process protections (U.S. Const. amend. V; U.S. Const. amend. VI). Implicit racial bias, as explained in Chapter One, not only can impact sentencing outcomes, especially through associations between race and criminality, but also can frame how information is interpreted and evaluated. This latter fact means that in addition to decreasing the fairness of the sentencing phase, implicit racial biases can also negatively influence other roles that the judge takes on throughout the trial, such as deciding what evidence to admit, whether to grant objections, and so forth. Now, it would be a different matter if there were no way to attenuate these implicit racial biases, as they might be accepted as a permanent flaw of human nature. Yet, if the neurointerventions do offer an effective means to ameliorate the issue, and therefore help meet the protections that the Bill of Rights guarantees, then by refusing to participate in the neurointerventions, judges would be jeopardizing a defendant's right to a fair trial. Moreover, since the trial system as a whole relies on promoting fairness, due process, and the presumption of innocence, a judge might not just have a moral duty to the defendant to partake in the neurointerventions, but also a moral duty to the justice system as an institution. Put differently, whereas the voluntary and pseudo-compulsory policies could justifiably provide significant benefits to defendants and the trial system, a judge's failure

to participate in the neurointerventions could be seen as inflicting harm to both parties. Thus, even though these policies entail forfeiting one's moral right to freedom of thought, judges should view neurointerventions as a personal obligation.

Importantly, the act of waiving rights to meet professional and moral duties is not unprecedented. Take the case of mandatory vaccinations in health care workers, which some states and hospitals currently require (National Center for Immunization and Respiratory Diseases, 2014). The comparison to judges can be made on multiple levels. First, health care workers choose to enter the profession and abide by the associated requirements, just as judges do. Second, the main role of health care workers is to provide medical assistance and, as articulated in medical ethics standards, to do no harm (Beauchamp & Childress, 2013). Contracting viruses that can spread to patients and negatively impact their health impedes a health care worker's ability to meet these professional obligations (Tilburt, Mueller, Ottenberg, Poland, & Koenig, 2008). Similarly, judges are expected to facilitate the administration of fair trials, and implicit racial biases hinder their capacity to execute this duty. Third, vaccines provide a potent means to reduce the chances of contracting viruses; neurointerventions, as this paper assumes will be the case in the future, also hold potential to be an effective attenuating factor. Fourth, in jurisdictions that have mandatory vaccination policies, individuals voluntarily waive their moral right to bodily integrity as a requirement of the health care profession, and allow a substance to be injected into their bodies that will have clear biological effects (Hooper, Breathnach, & Iqbal, 2014). Comparatively, judges would voluntarily waive their moral right to freedom of thought in order to meet professional

requirements, and participate in neurointerventions that have neurological effects. Fifth, health care workers in these jurisdictions do not just have a professional obligation to vaccinate themselves, but also a moral duty to not worsen their patient's health or the health of the general public (Antommara, 2013; Schwartz, 2013). Similarly, judges would have a professional duty to partake in the neurointerventions as well as obligations to respect the defendant's rights and protect the trial system's legitimacy. By voluntarily choosing to be in the health care profession, individuals take on duties that make participating in certain interventions (and waiving specific rights) ethically obligatory. Accordingly, there is precedence for claiming the existence of a moral imperative to voluntarily forfeit one's rights in a professional setting.

4.6 Neurointerventions and Judges: A Worthwhile Consideration

The prevalence of implicit racial biases in the majority of the general population is a well-documented and researched phenomenon (Banaji et al., 2003; Project Implicit, 2011). If that fact alone is not sufficiently troubling given the influence such biases can have on daily interactions, the presence of these biases in the courtroom setting is particularly problematic in light of the high stakes involved. As noted in the introductory chapter, what is perhaps most concerning of all is that the problems associated with implicit racial biases in the courtroom have been known for decades, and yet concerted and productive steps have not been taken to ameliorate the issue. This chapter has argued that should the reliability and efficacy of neurointervention techniques come to fruition, a policy asking or requiring judges to engage in the neurointerventions would not only be ethically justifiable, but also morally obligatory. While participating in the

neurointerventions would modulate a judge's freedom of thought, this chapter has asserted that a judge's autonomous decision to do so should be honored. Moreover, by entering into this line of work, judges agree to abide by professional requirements, impartially hear cases that come before them, uphold and respect the rights of the defendant, and contribute to the legitimacy of the trial system as a whole. As such, this paper has argued that the benefits of instituting neurointerventions as a professional requirement outweigh the potential consequences brought to those who choose not to participate and subsequently forfeit their eligibility to serve. Considering the pernicious influence that implicit racial biases can have on the quality and outcome of a trial, and that judges serve as the face of a theoretically impartial justice system, the prospect of instituting a neurointervention policy among judges should be viewed as a justifiable and worthwhile proposal.

5. Neurointerventions and Juries

Thus far, this paper has argued that the use of neurointerventions among judges should be regarded as a serious proposition. However, judges are not the only population in the courtroom setting affected by implicit racial biases. As mentioned in Chapter One, studies with mock juries have elucidated the strong influence that these biases can have on the interpretation of information and evaluation of guilt; consequently, reducing implicit racial biases in both judges and juries would serve to enhance the impartiality in a given trial.

The following chapter will take the framework used to analyze neurointerventions in judges and apply it to the context of jurors. Admittedly, the prospect of neurointerventions will be theoretical for both judges and jurors until more empirical data becomes available; however, in contrast to the argument with judges, the analysis for jurors is significantly more complicated given the unique characteristics and protections afforded to juries. Additionally, cultural acceptance of neurointervention policies might be especially difficult to achieve since jury service is not an optional duty, and logistical aspects might be more complex due to the size of juries for each trial (compared to a single judge). For these reasons, as well as others, the successful implementation of neurointervention policies in jurors will require overcoming a fair amount of barriers. Nevertheless, given the presence and impact of implicit racial biases among jurors, it is worth identifying and analyzing the relevant ethical and legal considerations.

5.1 Neurointervention Policies in Juries

As with the case of judges, implementing neurointerventions with jurors could be voluntary or pseudo-compulsory. A voluntary policy might take the following form. Upon being summoned for jury service, potential jurors would be told about the problem of implicit biases in the courtroom and given facts about the neurointerventions (including their safety and efficacy). If selected to serve, jurors would then have the option of participating in the neurointerventions prior to the start of the trial. The rationale for this proposed timing is to avoid the topic of neurointerventions from entering into the voir dire process. Otherwise, if the willingness to consent was a subject during jury selection, potential jurors might be swayed to answer in a manner counter to their actual preferences. Returning to the social desirability effects mentioned in Chapter One, potential jurors might give their consent when asked in the presence of the attorneys and judge in order to conform to what they believe is expected from them. Alternatively, one could also imagine that potential jurors might decline to give their consent when asked during voir dire solely as a strategy to get out of jury service. By waiting until after the jury panel has been selected, and conducting the consent process outside of the courtroom itself, some of these issues might be attenuated.

A pseudo-compulsory policy, on the other hand, would involve making the neurointerventions a requisite aspect of jury service. As such, in order to only conduct voir dire with those who could actually meet the requirements of the juror role, potential jurors would be informed about implicit biases and given the option to consent to the neurointerventions prior to the jury selection process. Only those jurors who agreed to the

neurointerventions would continue with voir dire, and if selected would partake in the neurointerventions before the trial commenced. However, to prevent the neurointerventions from simply providing an easy excuse to avoid serving, individuals who declined to give their consent would be required to fulfill some other civic duty.¹

Now that policies in the jury context have been articulated, let us address the ethical and legal issues that might arise.

5.2 The Ethics of Voluntary Neurointerventions

Given the nature of a voluntary policy, the ethical considerations for jurors are largely equivalent to those articulated for judges. In both cases, the individual chooses to partake in the neurointerventions without any resulting consequences for declining to do so.

Accordingly, a voluntary policy would uphold all four bioethical tenets. First, as previously argued, if individuals voluntarily choose to have their own freedom of thought modulated, and do not make this decision under coercive conditions, then the policy would be in compliance with the respect for autonomy. Second, potential harms to the juror would be minimized, assuming the neurointerventions were proven to be sufficiently safe and limited in scope prior to their implementation in the courtroom.

Third, since defendants are entitled by the Sixth Amendment to an impartial jury, participating in the neurointerventions would enable jurors to better fulfill their duties. As argued with the case of judges, the ability to more adequately meet expected standards might provide the individual juror with personal satisfaction and an increased sense of

¹ The exact nature of this civic duty is beyond the scope of this paper, but some safeguard would be necessary to prevent neurointerventions from becoming a pretense for not serving for reasons separate from the neurointerventions themselves.

competency in his role. By enabling jurors to better satisfy the demands of jury service, the voluntary policy would also increase a defendant's opportunity to have a fairer trial, and would legitimize the justice system as a whole by helping it satisfy its intended goals. Additionally, all citizens could benefit from this policy, since anyone could theoretically become a defendant at some point in the future. Fourth, since each juror would be given the same option to partake in the neurointerventions, and because the policy would not intentionally impose disproportionate burdens on only a specific subset of jurors, both comparative and distributive justice would be met. With respect to the policy's effects, reducing implicit racial biases in jurors would increase the chance that two cases involving comparative facts but defendants of differing races would be treated similarly. Moreover, since jury pools are randomly drawn (28 U.S.C. §1864), and each juror would have the same access to the neurointerventions, all defendants would have an equal opportunity to have their case heard by less biased jurors.

On a separate note, one potential concern with a voluntary policy in juries is that if the neurointerventions take place after jury selection, those chosen for certain reasons might behave differently during the trial than an attorney might have expected from the voir dire process. In contrast to judges, the identity or characteristics of those making up the jury panel is not a predetermined factor outside of an attorney's control. Instead, attorneys make intentional decisions during the voir dire process in choosing when to use their limited number of peremptory challenges (Vidmar & Hans, 2007). Attorneys therefore might strategically attempt to assemble a panel that will be most sympathetic to their case and use their challenges accordingly. One might claim that using

neurointerventions after voir dire could undermine the whole process of jury selection, as implicit biases that might have been expected to work for or against the defendant are no longer as present or influential. However, there are a few responses to this objection. First, if the ultimate purpose of jury selection is to exclude those from serving who cannot be impartial, then reducing implicit racial biases in jurors would support the goals of jury selection rather than undercut them. Second, even if attorneys purposefully base their challenges on a potential juror's implicit biases, such a strategy is grounded in prediction given that implicit biases are often unknown to the juror himself and can differ from his explicit beliefs. Therefore, even without neurointerventions, implicit biases may play an unexpected role. Third, defendants are entitled to an impartial jury of their peers (U.S. Const. amend. VI), not one with specific characteristics that are strategically desired by attorneys. As long as the neurointerventions specifically targeted implicit racial biases and did not have collateral impacts on fundamental aspects of a juror's personality or beliefs, the concern with voir dire does not appear especially problematic.

Putting this objection aside, this paper contends that a voluntary policy in jurors would satisfy the requirements of autonomy, nonmaleficence, beneficence, and justice. As such, the volitional use of neurointerventions among jurors should be regarded as ethically permissible.

5.3 The Ethics of Pseudo-Compulsory Neurointerventions

In contrast to the analysis of a voluntary policy, the considerations involved in a pseudo-compulsory policy diverge significantly from the case of judges. Before addressing the bioethical principles at stake, let us first consider the issue of a

representative jury. As a reminder, the pseudo-compulsory policy would involve only allowing those who consented to the neurointerventions to go through the voir dire process. One might wonder whether there is a certain type of individual who would agree to neurointerventions; in other words, those likely to provide consent might possess some common characteristic that makes them different from what one would find in a cross-section of the local community. For example, one concern might involve age, since people from younger generations tend to be more accepting of new technologies. At the same time, individuals from older generations might feel a stronger impetus to fulfill their civic duty and participate on a jury. Given that many factors will likely guide individuals' acceptance of neurointerventions and motivations to participate in them, it seems reasonable to expect that a fair cross-section of the community would still be represented. However, should future research show otherwise, the nature of the pseudo-compulsory policy would need to be adjusted. Let us now move on to examine the degree to which the pseudo-compulsory policy conforms to central bioethical principles.

5.3.1 Autonomy

While concerns with freedom of thought remain the same as in the voluntary policy (as well as in both policies for judges), the issue of coercion provokes further analysis. As noted in the previous chapter, coercion typically involves three main elements: (1) a reduced choice set, (2) a detrimental impact on the person making the decision, and (3) manipulative intent by the agent creating the options. This section will proceed by analyzing the relevant considerations for each point.

At a preliminary level, one could argue that the choice set of jurors has not been reduced, but merely modulated with additional conditions. In the existing system, if selected for the jury, individuals essentially have two choices—serve or face consequences of monetary fines or potential arrest. With the proposed policy, jurors have a more complicated set of options. A potential juror under the pseudo-compulsory scheme could choose to (a) consent to the neurointerventions and serve if selected, (b) consent to the neurointerventions, refuse to serve if selected, and subsequently face the same punitive ramifications as in the existing system, (c) not provide consent to the neurointerventions, forfeit the opportunity to serve if selected, and fulfill some alternative civic duty instead, or (d) not provide consent to the neurointerventions, forfeit the opportunity to serve, refuse to fulfill the alternative civic duty, and face punitive consequences. However, at the same time, one could also argue that a potential juror under the proposed policy no longer has the option of serving without participating in the neurointerventions, and thus an option has been taken away from him.

If we think that the option set has been reduced, the question of whether it has been reduced in an inherently detrimental way depends on whether a juror has a right to serve. Although there are a few legal provisions that come close, there does not appear to be an explicit legal right to jury service. To start, all citizens have a right to “have the opportunity to be considered for service” (28 U.S.C. §1861). However the right to be considered is not equivalent to the right to actually serve, and the opportunity to be summoned (which this provision seems to encapsulate) is different than the opportunity to be selected. Next, while the Supreme Court has acknowledged a right to not be

excluded from service for reasons such as race (*Powers v. Ohio*, 1991), a positive right to serve has not been officially pronounced.² Moreover, the Supreme Court has recognized states' abilities to set "relevant qualifications" for juror eligibility (*Carter v. Jury Commission of Greene County*, 1970). The willingness to participate in the neurointerventions could plausibly be considered a relevant factor, since impartiality is a core tenet of the jury system. Nevertheless, some scholars assert that even though there is not an explicitly stated right to serve, this right is implicit in other rights, such as the right to vote, or First Amendment rights protecting civic participation (Amar & Hirsch, 1999; Ferguson, 2013). Whether or not these arguments actually establish a legal right to serve, they seem to imply a moral right or privilege to serve. Jury service allows citizens to contribute to the societal and legal framework under which they operate on a daily basis and uphold fundamental Constitutional values that comprise a democratic nation; citizens might therefore be morally entitled to have the opportunity to serve by being members of this society. If there is in fact a moral right to serve, then according to the reasoning employed in the previous chapter, the creation of a forced choice set that violates this right would reduce a juror's welfare and constitute a threat. In other words, one could claim that by making jurors forfeit their moral right to serve unless they participate in the neurointerventions, such a policy would be detrimental to the juror's wellbeing.

The third prong of the coercion definition, manipulation, is more difficult to argue since as mentioned with the case of judges, the intent of a pseudo-compulsory policy is to

² Interestingly, an American Bar Association paper cites *Powers v. Ohio* (1991) as articulating a legal right of all citizens to serve, despite the fact that the actual opinion does not support such a claim (American Bar Association Commission on the American Jury Project, 2005, p. 11).

increase the fairness of the trial system, not to make jurors do something against their will. However, given the fact that jury service is mandatory if an individual is selected, and that citizens might have a strong desire to participate in the justice system, one could argue that the government, by instituting a pseudo-compulsory policy, is exploiting the interests of jurors to advance its own goals.

Despite the fact that claims can be made on either side of the spectrum, it appears that a case can plausibly be argued for considering a pseudo-compulsory policy coercive. This evaluation does not mean, however, that all individuals would necessarily be coerced into agreeing to the neurointerventions. For instance, some individuals might not have an interest in serving. Given that many citizens try to get out of jury service and view it as an unpleasant burden (Amar & Hirsch, 1999), the idea that some individuals would uninhibitedly choose to waive their moral right to service seems reasonable or even predictable. Additionally, other individuals might actively desire to participate in the neurointerventions, and thus would not be pressured into making the decision to do so. That being said, the consent of individuals who harbor a strong interest in serving and possess qualms about the neurointerventions might indeed be undermined by the coercive nature of the policy.

5.3.2 Nonmaleficence

Assuming the neurointerventions were adequately safe and effective, one potential concern might involve the negative consequences that could occur for the individual who declines to provide his consent and subsequently forfeits his ability to serve. Although many people in the jury pool are not selected to serve despite being eligible, surrendering

the mere opportunity to be selected might be perceived as a barrier to public engagement or an affront to one's contributions as a functioning member of society. Additionally, for those with a strong interest in serving who felt coerced into providing their consent, psychological distress might result from participating in the neurointerventions despite their preferred desire not to do so.

5.3.3 Beneficence

As discussed with the voluntary policy, the neurointerventions would provide benefits to the individual juror, the trial system, defendants, and all members of society (since anyone could find themselves in the position of a defendant). Additionally, while having just one less biased juror on a trial might be better than none, having all jurors on a trial be less biased is likely to facilitate fairer outcomes. In other words, a pseudo-compulsory policy would guarantee more substantial benefits to the trial system than a voluntary policy, since all members of all juries would have reduced implicit racial biases. Similarly, defendants are entitled to an impartial jury, as opposed to simply a panel with a few impartial individuals; although the neurointerventions would not eliminate all implicit racial biases, let alone ensure complete impartiality, reducing implicit racial biases among all members of the jury has the potential to respect the defendant's Sixth Amendment rights better than either a voluntary policy or the status quo.

5.3.4 Justice

With respect to comparative and distributive justice, providing all potential jurors with the same opportunity and requirement to participate in the neurointerventions would ensure that each juror is treated in a similar manner. Moreover, this requirement would

distribute the burdens more equally across jurors (and perhaps across the citizenry in general given that jury service is a mandatory civic duty). As a result, a pseudo-compulsory policy would alleviate concerns with free-riding, since all jurors would be contributing to the public benefits of the policy instead of only a select few.

Since every criminal case would be heard by a jury with reduced implicit racial biases, the chance of white and non-white defendants with similar cases receiving similar treatment (especially in the form of access to a fair and impartial trial) would presumably be increased. Furthermore, making neurointerventions a requisite aspect of jury service would allow each defendant to derive benefits from the policy.

5.3.5 The Ethical Permissibility of Pseudo-Compulsory Neurointerventions

Compared to the other policies examined thus far, the pseudo-compulsory system in jurors appears to present the most potential concerns. For one, the autonomy of some jurors might be breached given the coercive nature of the policy. As a result, the policy might cause psychological harm to the jurors who feel coerced to consent despite their contrary desires, as well as interfere with the interests of individuals who decline to provide consent and surrender their eligibility. At the same time, the pseudo-compulsory policy would provide substantial benefits to defendants, the justice system, and society at large by facilitating more impartial trials in a justly distributed manner. The policy additionally would help jurors to meet the expectations of impartiality that are demanded of them.

The principle of beneficence requires that the risks of harm to the individual must be outweighed by the benefits. As stated in the discussion of the pseudo-compulsory policy

in judges, the harm brought to the individual who declines to consent and surrenders his eligibility does not seem as problematic when we consider the fundamental aspects of the juror role. Jurors are required to impartially hear the case before them, and acting in accordance with the expectations of the role serves to uphold the legitimacy of the trial system. Thus, if an individual has a substantial interest in engaging in the justice system through this civic duty, then the policy should be viewed in a favorable light since it advances the goals that jury service entails.

As such, this paper asserts that the more pressing ethical issue is the threat to a juror's autonomy. However, autonomy is not an absolute right, especially in situations that inherently involve the welfare of others (Oshana, 2003). Accordingly, overriding a juror's right to autonomy could be justified given compelling government interests regarding the larger society. For instance, John Stuart Mill's harm principle contends that government restriction of individual liberties is only justified when not doing so would cause harm to others (as cited in Stanton-Ife, 2009). A juror's failure to participate in the neurointerventions could constitute harm in at least two ways. First, maintaining implicit racial biases could directly harm the defendant by reducing the overall fairness of the trial and hindering his Sixth Amendment right to an impartial jury. If implicit racial biases prime jurors to make automatic associations between race and criminality, the defendant's right to the presumption of innocence would be compromised. Second, one could argue that such a refusal would also constitute a public harm (or a harm to public interests) since it would decrease the fairness and legitimacy of the trial system as a whole, undermining a fundamental institution upon which the society depends (as cited in

Stanton-Ife, 2009). Therefore, given that the pseudo-compulsory policy would prevent harm to defendants and the system, enable trials to better meet the standards guaranteed by the Sixth Amendment, and distribute benefits across all criminal proceedings, this paper argues that a juror's autonomy and interests can be justifiably overridden.

The arguments put forth in this section might appear to unduly prioritize societal benefits at the expense of the individual. It is acknowledged that ethical theories centered on respecting autonomy and nonmaleficence might not support the justifications posited above, and this paper does not seek to overlook the importance of individual rights. However, it is essential to remember that jury duty is a unique role, and its mandatory nature could be said to undermine a juror's autonomy from the start, as it ignores whether an individual actually desires to perform the task. Additionally, jury service is not a phenomenon that involves self-regarding or isolated actions. Instead, jury duty in its very essence is a service to society, and members of the community are charged with evaluating whether the defendant breached societal rules. Furthermore, as previously stated, all citizens are subject to these rules and standards, and could find themselves in the position of a defendant at some point in the future. As such, contributing to society could confer direct benefits to the individual juror (besides those already articulated pertaining to more adequate fulfillment of duties). In sum, jury duty serves a societal function by definition, and the central role that the justice system plays in society affects all citizens through their membership in the community. As a result, this chapter argues that within the context of jury duty, the principles of beneficence and justice can justifiably outweigh concerns with autonomy and nonmaleficence. This paper therefore

contends that a pseudo-compulsory neurointervention policy in jurors would be ethically permissible.

5.4 Neurointerventions as a Moral Obligation

This chapter has argued that both a voluntary and pseudo-compulsory neurointervention policy would be ethically permissible to implement. However, even if we are more comfortable with a voluntary policy than a pseudo-compulsory one (given the concerns addressed above), the following section will maintain that jurors should regard neurointerventions as a moral obligation.

Whereas the law is silent on the question of whether jury service is a legal right, it is unequivocal in positioning jury service as a mandatory civic duty. Importantly, the law already has a set of standards in place to help enable juror impartiality and prevent those deemed incapable of fulfilling their duties from serving. For instance, individuals considered too young, with insufficient English language fluency, with certain mental disabilities, or with obvious conflicts of interest are all ineligible for service (28 U.S.C. §1865). Moreover, during the period of a trial, jurors who do serve are prohibited from speaking to each other at specific times about certain topics, their freedom of movement may be restricted, and their access to various sources of information may be limited (Judicial Conference of the United States, n.d.). These regulations reflect an expectation that jurors not only perform their civic duty, but also perform it well.

Given that implicit racial biases can negatively influence the capacity for impartiality, such biases would impede a juror's ability to adequately fulfill his civic duty. If neurointerventions were made available, and provided a means to effectively ameliorate

implicit racial biases, it seems reasonable to find their use more than just ethically justifiable, but actually morally obligatory. Not only would the failure to alleviate these biases prevent jurors from meeting their own moral and civic responsibilities, but as previously mentioned, such a refusal would also cause harm to the defendant (by infringing upon his Constitutional right to an impartial jury), the trial system (by diminishing its legitimacy or ability to achieve its intended purpose), and society at large (by undermining its commitment to a fair justice process).

5.5 Neurointerventions in Juries: Conclusions from a Conceptual Analysis

The system of trial by jury affords citizens with a remarkably important role in society. Individuals eligible for service are tasked with appearing for jury duty if summoned, devoting their time and energy to a case if selected, and ultimately deciding on matters that could completely change the course of another citizen's life. While jury duty is mandatory for those eligible, it is also an opportunity for citizens to partake in a democratic process and promote the values and liberties central to the Constitution. The fact that jurors have certain obligations as a function of this civic duty, but at the same time do not actively choose to take on this role, makes for a complex theoretical analysis regarding neurointerventions. While recognizing the relevant issues with coercion and nonmaleficence, this chapter has argued that using neurointerventions with jurors could not only be viewed as ethically justifiable, but also morally obligatory. Juries are considered a special and protected entity given their important role in administering justice. However, if the jury is such a vital aspect of the justice system and is crucial for the preservation of a democratic society, then we also have to acknowledge that jury

service is a duty that entails certain requirements. If jurors are not impartial, they cannot adequately uphold the values of fairness and the presumption of innocence that the justice system and democracy demand. In other words, if we claim that citizens have a moral right to jury service, then it seems reasonable to expect them to thoroughly fulfill the obligations that accompany the role.

As discussed earlier, the prospect of using neurointerventions in jurors is complicated by the unique and mandatory nature of jury duty. Moreover, significant practical and logistical barriers might arise due to the size of juries and the impact that individual differences might have on the efficacy of the neurointerventions. Accordingly, the actual implementation of neurointervention policies in juries would likely be difficult to achieve. Nevertheless, the concept of using neurointerventions with jurors at the very least provokes discussion as to the role of jurors, the standards and obligations they are intended to meet, and the extent to which they should be expected to fulfill such duties.

6. Neurointerventions in the Courtroom: Final Remarks and Future Directions

6.1 Summary

Our criminal justice system places substantial value on fairness and the presumption of innocence. As a result, a plethora of rules and protections exist to prevent the conviction of an innocent individual, even at the cost of potentially acquitting a guilty perpetrator. For instance, the burden of proof is on the prosecution, the threshold for a guilty verdict is beyond a reasonable doubt, defendants are guaranteed the right to a fair trial by an impartial jury, and Constitutional amendments emphasize due process during trials. While this framework is revered for promoting a transparent and just system, the prevalence and impact of implicit racial biases in the courtroom serve to undermine the realization of these values and standards. If black defendants tend to be automatically associated with guilt, criminality, and threat on a subconscious level, especially by those in decision-making positions such as judges and jurors, it becomes impossible to truly achieve the requirements of impartiality and the presumption of innocence.

The problem with implicit racial bias in the courtroom extends beyond this inability to meet expected standards. For one, biased judgments can have far reaching ramifications, as decisions made in the criminal justice system can dictate matters of life or death, as well as deprive individuals of their freedom. Additionally, the stressful, uncertain, and cognitively demanding environment of the courtroom setting combines elements known to exacerbate implicit biases. As a result of these aggravating conditions, the strategies commonly advocated for decreasing implicit racial biases are insufficient.

In light of the persisting impact of implicit racial biases in the courtroom, this paper has put forth an unconventional and novel approach in the search for a solution. While much research remains to be conducted, neurointerventions have the potential to reduce implicit racial biases more effectively than the measures proposed in the existing literature. These neurointerventions might take the form of computerized brain-training tasks or noninvasive brain stimulation techniques, and tailored research efforts will be necessary to discover and develop the most feasible and effective mechanisms. Nevertheless, by reducing automatic self-outgroup distinctions and enabling better self-regulatory capacities, neurointerventions could plausibly become a useful mitigation strategy for the courtroom setting, especially in the case of judges. One significant objection to such an approach is the interference with a judge's freedom of thought. While a legitimate concern, if judges choose to voluntarily surrender their freedom of thought, then respecting autonomy necessitates honoring this decision. Moreover, including neurointerventions as a part of judges' professional obligations should neither be viewed as coercive nor ethically impermissible. While a judge might be less well off if he decides not to participate in the neurointerventions (and thereby forfeits his ability to serve as a judge), this potential negative consequence must be balanced against the fact that judges elect to enter into the field and are not entitled to retain their position if they do not meet the required standards. In fact, given that judges are tasked with upholding a defendant's right to a fair trial, and that a judge's own implicit racial bias would compromise such a right, judges can reasonably be said to have a moral imperative to participate in the neurointerventions to fulfill their judicial duties. Thus, this paper has

made three major claims: (1) that neurointerventions (such as computerized brain-training tasks and noninvasive brain stimulation techniques) could comprise effective de-biasing solutions in the future, (2) that the use of neurointerventions in judges would be ethically justified, and (3) that judges, as a function of their position and associated duties, have a moral obligation to participate in the neurointerventions. Lastly, this paper analyzed the use of neurointerventions in jurors through a more theoretical lens, articulating the main ethical and legal divergences from the case of judges. With a pseudo-compulsory situation in particular, prominent concerns include (1) the coercive nature of the policy, and (2) the deprivation of a moral right to jury service for those who choose not to participate in the neurointerventions. However, this paper has argued that the benefits of the policy outweigh these burdens, especially considering the fact that jury duty is a service to society, and any individual selected to serve should be expected to adequately satisfy the demands of the role. Therefore, this paper contends that the use of neurointerventions with jurors could be viewed as ethically justifiable and morally obligatory as well.

6.2 Future Directions

While the use of neurointerventions among jurors might remain in the theoretical realm due to practical barriers, the implementation of such a mitigation strategy in judges should be considered a serious proposition. This is not to say that judges would readily embrace the policy; as discussed in Chapter One, the bias blindspot and the social desirability effect make it exceedingly difficult to accept or admit to one's own biases, particularly in individuals who hold a position explicitly exalted for its impartial nature.

Nevertheless, the acceptance of a neurointervention policy is not unobtainable, especially if incremental steps are taken. For instance, usage could be strictly voluntary during initial implementation phases in order to expose judges to the mechanisms and allow them to gain familiarity with the practice. Should multiple methods prove to be sufficiently effective, the next step might be to introduce a pseudo-compulsory policy allowing judges to choose their preferred neurointervention technique, and perhaps eventually transition into offering only the most efficacious options. Another potential catalyst for acceptance is the way in which the policies are framed. For example, the neurointerventions could be explained as a training technique that better enables judges to fulfill their roles. Accurately describing the neurointerventions as a tool, rather than a scientific solution to reduce human errors, would help to position the mechanisms as useful methods for judges to incorporate into their preexisting practices as opposed to a blatant expression of judges' shortcomings.

Although the potential neurointerventions proposed in Chapter Three are not ready to be implemented in the courtroom, it is important to note that some crucial groundwork has already been laid. To start, computerized brain-training games, virtual reality devices, and tDCS are all increasing in demand and availability in the general population (Jwa, 2015; Poon, 2015). By the time that sufficient research has been conducted for their use in the courtroom setting, these techniques will likely be regarded as relatively familiar phenomena, a factor which might help reduce initial aversions to the technologies. Second, the presence and influence of implicit racial biases are not just becoming a more discussed issue among scholars, but also among the general public. Within the past year,

there has been a surge of newspaper articles, blog posts, and other media forums calling attention to the problem and expressing a need for change (Mullainathan, 2015; Noë, 2015). Third, the translation of increased awareness into systemic reform has already begun to take shape in other aspects of the justice system. For instance, discrepancies in the setting of bail across similar cases have revealed problems with implicit racial bias, the limitations of judicial discretion, and a failure of common practice to meet its intended purpose. In response, multiple jurisdictions have begun to implement an algorithmic measure shown to better predict the factors upon which bail is supposed to be based (Dewan, 2015). While initial acceptance of the measure was met with resistance, those jurisdictions that have integrated it into the decision-making process have become more confident in the increased fairness and legitimacy of the justice system. Moreover, similar to the framework suggested for neurointerventions, the algorithm has been promoted as a supplemental aid to decisions rather than a stand-alone solution that encroaches upon the judge's role (Dewan, 2015). Together, these recent developments suggest that a neurointervention policy to reduce implicit racial biases is not purely hypothetical: the technologies are becoming more mainstream, the general public is aware and concerned about the issue of implicit bias, and steps towards systemic changes in the courtroom have been set in motion.

While studies on neurointerventions have not yet been conducted in the courtroom setting, the research findings we do have suggest that certain techniques (such as computerized brain-training tasks or noninvasive brain stimulation) could be promising and feasible strategies for reducing implicit racial bias. However, even if these

mechanisms (or similar ones in the future) do not live up to these expectations, the discussion of neurointerventions in the courtroom is still a fruitful endeavor. If nothing else, such a proposition should inspire further conversation about the use of innovative methods for curbing implicit racial biases. The justice system was designed to ensure fair and impartial trials based on a presumption of innocence, and the pernicious influence of implicit racial biases impedes the realization of these goals as well as jeopardizes the justice system's legitimacy. Given advances in brain and behavioral sciences, implicit racial biases should not be considered an immutable flaw of human nature. In the past decade, researchers have made immense progress in elucidating neural regions and processes that contribute to the presence, influence, and regulation of these biases. If we want defendants to enjoy the protections afforded to them by the Bill of Rights, we should harness this knowledge and utilize it to address a problem that undermines one of the most fundamental systems in our society. Neurointerventions might not be the ultimate answer, but they deserve our consideration.

References

28 U.S.C. § 144

28 U.S.C. § 455

28 U.S.C. § 1861

28 U.S.C. § 1864

28 U.S.C. § 1865

Ahn, S. J., Bailenson, J. N., & Park, D. (2014). Short- and long-term effects of embodied experiences in immersive virtual environments on environmental locus of control and behavior. *Computers in Human Behavior*, *39*, 235-245. doi: 10.1016/j.chb.2014.07.025

Amar, A. R., & Hirsch, A. R. (1999). *For the people: What the constitution really says about your rights*. New York, NY: Simon & Schuster.

American Bar Association Commission on the American Jury Project. (2005). *Principles for juries and jury trials*. Retrieved from http://www.americanbar.org/content/dam/aba/migrated/2011_build/american_jury/final_commentary_july_1205.authcheckdam.pdf

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670-682. doi: 10.1038/nrn3800

Antommaria, A. H. (2013). An ethical analysis of mandatory influenza vaccination of health care personnel: Implementing fairly and balancing benefits and burdens. *American Journal of Bioethics*, *13*(9), 30-37. doi: 10.1080/15265161.2013.814731

Banaji, M. R., Bazerman, M. H., & Chugh, D. (2003, December). How unethical are you? *Harvard Business Review*, 1-9. Retrieved from <https://hbr.org/2003/12/how-unethical-are-you>

Banissy, M. J., & Ward, J. (2013). Mechanisms of self-other representations and vicarious experiences of touch in mirror-touch synesthesia. *Frontiers in Human Neuroscience*, *7*, 1-3. doi: 10.3389/fnhum.2013.00112

Beauchamp, T. L., & Childress, J. F. (2013). *Principles of biomedical ethics* (7th Ed.). New York, NY: Oxford University Press.

- Blair, I. V., Judd, C. M., & Chapleau, K. M. (2004). The influence of Afrocentric facial features in criminal sentencing. *Psychological Science, 15*(10), 674-679. doi: 10.1111/j.0956-7976.2004.00739.x
- Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*(5), 828-841. doi: 10.1037//0022-3514.81.5.828
- Brunoni, A. R., Nitsche, M. A., Bolognini, N., Bikson, M., Wagner, T., Merabet, L., . . . Fregni, F. (2012). Clinical research with transcranial direct current stimulation (tDCS): Challenges and future directions. *Brain Stimulation, 5*(3), 175-195. doi: 10.1016/j.brs.2011.03.002
- Bublitz, C. (2015). Moral enhancement and mental freedom. *Journal of Applied Philosophy, 32*(3), 1-19. doi: 10.1111/japp.12108
- Bublitz, J. C., & Merkel, R. (2012). Crimes against minds: On mental manipulations, harms and a human right to mental self-determination. *Criminal Law and Philosophy, 8*(1), 51-77. doi: 10.1007/s11572-012-9172-y
- Bull, R. (1982). Physical appearance and criminality. *Current Psychological Reviews, 2*, 269-282. Retrieved from <http://link.springer.com/article/10.1007%2FBF02684461>
- Carter v. Jury Commission of Greene County, 396 U.S. 320 (1970).
- Cattaneo, Z., Mattavelli, G., Platania, E., & Papagno, C. (2011). The role of the prefrontal cortex in controlling gender-stereotypical associations: A TMS investigation. *Neuroimage, 56*(3), 1839-1846. doi: 10.1016/j.neuroimage.2011.02.037
- Caviola, L., & Faber, N. S. (2014). How stress influences our morality. *In-Mind Magazine*. Retrieved from <http://www.in-mind.org/article/how-stress-influences-our-morality>
- Coffman, B. A., Clark, V. P., & Parasuraman, R. (2014). Battery powered thought: Enhancement of attention, learning, and memory in healthy adults using transcranial direct current stimulation. *Neuroimage, 85*(3), 895-908. doi: 10.1016/j.neuroimage.2013.07.083
- Congressional Research Service. (2014). *The Constitution of the United States of America: Analysis and interpretation* (Senate Document No. 112-9, 2nd Sess). Retrieved from <http://www.gpo.gov/fdsys/pkg/GPO-CONAN-REV-2014/pdf/GPO-CONAN-REV-2014.pdf>

- Corbyn, Z. (2011). Hungry judges dispense rough justice. *Nature News*. doi: 10.1038/news.2011.227
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83(6), 1314-1329. doi: 10.1037/0022-3514.83.6.1314
- Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, 108(17), 6889-6892. doi: 10.1073/pnas.1018033108
- Dasgupta, N., Desteno, D., Williams, L. A., & Hunsinger, M. (2009). Fanning the flames of prejudice: The influence of specific incidental emotions on implicit prejudice. *Emotion*, 9(4), 585-591. doi: 10.1037/a0015961
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81(5), 800-814. doi: 10.1037/0022-3514.81.5.800
- DeGrassi, S. W., Morgan, W. B., Walker, S. S., Wang, Y., & Sabat, I. (2012). Ethical decision-making: Group diversity holds the key. *Journal of Leadership, Accountability and Ethics*, 9(6), 51-65. Retrieved from http://m.www.na-businesspress.com/JLAE/DeGrassiSW_Web9_6_.pdf
- Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental Social Psychology*, 48(6), 1267-1278. doi: 10.1016/j.jesp.2012.06.003
- Dewan, S. (2015, June 26). Judges replacing conjecture with formula for bail. *The New York Times*. Retrieved from <http://www.nytimes.com/2015/06/27/us/turning-the-granting-of-bail-into-a-science.html?emc=eta1>
- Douglas, T. (2013). Moral enhancement via direct emotion modulation: A reply to John Harris. *Bioethics*, 27(3), 160-168. doi: 10.1111/j.1467-8519.2011.01919.x
- Falcone, B., Coffman, B. A., Clark, V. P., & Parasuraman, R. (2012). Transcranial direct current stimulation augments perceptual sensitivity and 24-hour retention in a complex threat detection task. *PLoS ONE*, 7(4), e34993. doi: 10.1371/journal.pone.0034993

- Farah, M. J., Smith, M. E., Ilieva, I., & Hamilton, R. H. (2014). Cognitive enhancement. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1), 95-103. doi: 10.1002/wcs.1250
- Fed. R. Crim. P. 24
- Feinberg, J. (1986). Autonomy. In *The moral limits of the criminal law: Harm to self* (Vol. 3, pp. 27-51). New York, NY: Oxford University Press.
- Ferguson, A. G. (2013). *Why jury duty matters: A citizen's guide to constitutional action*. New York, NY: NYU Press.
- Fox, M. D., Halko, M. A., Eldaeif, M. C., & Pascual-Leone, A. (2012). Measuring and manipulating brain connectivity with resting state functional connectivity magnetic resonance imaging (fcMRI) and transcranial magnetic stimulation (TMS). *Neuroimage*, 62(4), 2232-2243. doi: 10.1016/j.neuroimage.2012.03.035
- Freeman, J. B., Penner, A. M., Saperstein, A., Scheutz, M., & Ambady, N. (2011). Looking the part: Social status cues shape race perception. *PLoS ONE*, 6(9), e25107. doi: 10.1371/journal.pone.0025107
- Gallate, J., Wong, C., Ellwood, S., Chi, R., & Snyder, A. (2011). Noninvasive brain stimulation reduces prejudice scores on an implicit association test. *Neuropsychology*, 25(2), 185-192. doi: 10.1037/a0021102
- Geyh, C.G. (2010) *Judicial disqualification: An analysis of federal law* (2nd ed.). Washington, DC: Federal Judicial Center. Retrieved from [http://www.fjc.gov/public/pdf.nsf/lookup/judicialdq.pdf/\\$file/judicialdq.pdf](http://www.fjc.gov/public/pdf.nsf/lookup/judicialdq.pdf/$file/judicialdq.pdf)
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for Black and White patients. *Journal of General Internal Medicine*, 22(9), 1231-1238. doi: 10.1007/s11606-007-0258-5
- Greene, E., & Heilbrun, K. (2011). *Wrightsmen's psychology and the legal system* (7th ed.). Belmont, CA: Wadsworth.
- Habel, P., & Scott, K. (2014). New measures of judges' caseload for the federal district courts, 1964–2012. *Journal of Law and Courts*, 2(1), 153-170. doi: 10.1086/670669
- Hawkins, J. S., & Emanuel, E. J. (2005). Confusions about coercion. *Hastings Center Report*, 35(2), 16-19. doi: 10.1353/hcr.2005.0075

- Hayes, J. R., Sheedy, J. E., Stelmack, J. A., & Heaney, C. A. (2007). Computer use, symptoms, and quality of life. *Optometry & Vision Science, 84*(8), E738-E755. doi: 10.1097/OPX.0b013e31812f7546
- Hooper, C. R., Breathnach, A., & Iqbal, R. (2014). Is there a case for mandating influenza vaccination in healthcare workers? *Anaesthesia, 69*(2), 95-100. doi: 10.1111/anae.12561
- Howe, S. W. (1995). Juror neutrality or an impartiality array? A structural theory of the impartial jury mandate. *Notre Dame Law Review, 70*(5), 1173-1245. Retrieved from <http://scholarship.law.nd.edu/ndlr/vol70/iss5/3/>
- Hulsken, S., Martin, A., Mohajeri, M. H., & Homberg, J. R. (2013). Food-derived serotonergic modulators: Effects on mood and cognition. *Nutrition Research Reviews, 26*(2), 223-234. doi: 10.1017/S0954422413000164
- iCivics. (n.d.). *Our story*. Retrieved from <https://www.icivics.org/our-story>
- Jacquet, P. O., & Avenanti, A. (2013). Perturbing the action observation network during perception and categorization of actions' goals and grips: State-dependency and virtual lesion TMS effects. *Cerebral Cortex, 25*(3), 598-608. doi: 10.1093/cercor/bht242
- Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology, 41*(3), 137-146. doi: 10.1027/1864-9335/a000020
- Judicial Conference of the United States. (n.d.). *Handbook for trial jurors serving in the United States district courts*. Washington, DC: Administration Office of the United States Courts. Retrieved from <http://www.txwd.uscourts.gov/Jury/Documents/handbookfortrialjurors.pdf>.
- Jwa, A. (2015). Early adopters of the magical thinking cap: A study on do-it-yourself (DIY) transcranial direct current stimulation (tDCS) user community. *Journal of Law and the Biosciences*. doi: 10.1093/jlb/lsv017
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Kang, J., Bennett, J. M., Carbado, D., Casey, P., Dasgupta, N., Faigman, D., . . . Mnookin, J. (2012). Implicit bias in the courtroom. *UCLA Law Review, 59*(5), 1124-1186. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=78176008&site=eds-live>

- Knäuper, B., Roseman, M., Johnson, P. J., & Krantz, L. H. (2009). Using mental imagery to enhance the effectiveness of implementation intentions. *Current Psychology*, 28(3), 181-186. doi: 10.1007/s12144-009-9055-0
- Larson, D. K. (2010). A fair and implicitly impartial jury: An argument for administering the implicit association test during voir dire. *DePaul Journal for Social Justice*, 3(2), 139-172. Retrieved from http://heinonline.org/HOL/Page?handle=hein.journals/depjsj3&g_sent=1&id=156
- Lawson, B. D. (2014). Motion sickness symptomology and origins. In K. S. Hale & K. M. Stanney (Eds.), *Handbook of virtual environments: Design, implementation, and applications* (2nd ed., pp. 531-600). Boca Raton, FL: CRC Press.
- Levine, R. J. (1988). *Ethics and regulation of clinical research* (2nd ed.). New Haven, CT: Yale University Press.
- Levinson, J. D., Cai, H., & Young, D. (2010). Guilty by implicit racial bias: The guilty/not guilty implicit association test. *Ohio State Journal of Criminal Law*, 8, 187-208. Retrieved from <http://ssrn.com/abstract=1471567>
- Levinson, J. D., & Young, D. (2010). Different shades of bias: Skin tone, implicit racial bias, and judgments of ambiguous evidence. *West Virginia Law Review*, 112, 307-350. Retrieved from <http://ssrn.com/abstract=1601615>
- Maister, L., Sebanz, N., Knoblich, G., & Tsakiris, M. (2013). Experiencing ownership over a dark-skinned body reduces implicit racial bias. *Cognition*, 128(2), 170-178. doi: 10.1016/j.cognition.2013.04.002
- Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: Owning another body affects social cognition. *Trends in Cognitive Science*, 19(1), 6-12. doi: 10.1016/j.tics.2014.11.001
- Mayo Clinic Staff. (2015, February 4). *Transcranial magnetic stimulation: Risks*. Retrieved from <http://www.mayoclinic.org/tests-procedures/transcranial-magnetic-stimulation/basics/risks/prc-20020555>
- Maznevski, M. L. (1994). Understanding our differences: Performance in decision-making groups with diverse members. *Human Relations*, 47(5), 531-552. doi: 10.1177/001872679404700504
- Melnick, M. (2011, April 14). When lunch is served, so is justice. *Time*. Retrieved from <http://healthland.time.com/2011/04/14/when-lunch-is-served-so-is-justice/>
- Mooney, C. (2014, December 8). Across America, Whites are biased and they don't even know it. *The Washington Post*. Retrieved from

<http://www.washingtonpost.com/blogs/wonkblog/wp/2014/12/08/across-america-whites-are-biased-and-they-dont-even-know-it/>

Mullainathan, S. (2015, January 3). Racial bias, even when we have good intentions. *The New York Times*. Retrieved from <http://www.nytimes.com/2015/01/04/upshot/the-measuring-sticks-of-racial-bias-.html?abt=0002&abg=0>

N.C. Gen. Stat. § 15A-1223

National Center for Immunization and Respiratory Diseases. (2014). *State immunization laws for healthcare workers and patients*. Retrieved from <http://www2a.cdc.gov/vaccines/statevaccsApp/AdministrationbyPatientType.aspx?PatientTypetmp=HospitalEmployees> - 86

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., . . . Pascual-Leone, A. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation*, 1(3), 206-223. doi: 10.1016/j.brs.2008.06.004

Noë, A. (2015, January 9). *The Biased Eye*. Retrieved from <http://www.npr.org/sections/13.7/2015/01/09/376039529/the-biased-eye>

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, 31(2), 166-180. doi: 10.1177/0146167204271418

Oshana, M. (2003). How much should we value autonomy? In E.F. Paul, F.D. Miller Jr., & J. Paul (Eds.), *Autonomy* (pp. 99-126). Cambridge: Cambridge University Press.

Paladino, M., Mazurega, M., Pavani, F., & Schubert, T. W. (2010). Synchronous multisensory stimulation blurs self-other boundaries. *Psychological Science*, 21(9), 1202-1207. doi: 10.1177/0956797610379234

Papillon, K. (2013). The court's brain: Neuroscience and judicial decision making in criminal sentencing. *Court Review*, 49(1), 48-62. Retrieved from http://heinonline.org/HOL/Page?handle=hein.journals/ctrev49&g_sent=1&id=48

Payne, K. B. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81(2), 181-192. doi: 10.1037/0022-3514.81.2.181

Payne, K. B. (2006). Weapon bias: Split-second decisions and unintended stereotyping. *Current Directions in Psychological Science*, 15(6), 287-291. doi: 10.1111/j.1467-8721.2006.00454.x

- Payne, K. B., Lambert, A. J., & Jacoby, L. L. (2002). Best laid plans: Effects of goals on accessibility bias and cognitive control in race-based misperceptions of weapons. *Journal of Experimental Social Psychology*, 38(4), 384-396. doi: 10.1016/S0022-1031(02)00006-9
- Peck, T. C., Seinfeld, S., Aglioti, S. M., & Slater, M. (2013). Putting yourself in the skin of a Black avatar reduces implicit racial bias. *Consciousness and Cognition*, 22(3), 779-787. doi: 10.1016/j.concog.2013.04.016
- Phillips, K. W., Northcraft, G. B., & Neale, M. A. (2006). Surface-level diversity and decision-making in groups: When does deep-level similarity help? *Group Processes & Intergroup Relations*, 9(4), 467-482. doi: 10.1177/1368430206067557
- Presidential Commission for the Study of Bioethical Issues. (2015). *Gray matters: Topics at the intersection of neuroscience, ethics, and society*. Retrieved from http://bioethics.gov/sites/default/files/GrayMatter_V2_508.pdf
- Poon, L. (2015, July 5). *Want a taste of virtual reality? Step one: Find some cardboard*. Retrieved from <http://www.npr.org/sections/goatsandsoda/2015/07/05/419543981/want-a-taste-of-virtual-reality-step-one-find-some-cardboard>
- Powers v. Ohio, 499 U.S. 400 (1991).
- Project Implicit. (2011). *Frequently asked questions*. Retrieved from <http://implicit.harvard.edu/implicit/faqs.html>
- Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does unconscious racial bias affect trial judges? *Notre Dame Law Review*, 84(3), 1195-1246. Retrieved from <http://ssrn.com/abstract=1374497>
- Reynolds, C. (2013). Implicit bias and the problem of certainty in the criminal standard of proof. *Law & Psychology Review*, 37, 229-248. Retrieved from http://heinonline.org/HOL/Page?handle=hein.journals/lpsyr37&g_sent=1&id=237
- Roberts, A. (2012). (Re)forming the jury: Detection and disinfection of implicit juror bias. *Connecticut Law Review*, 44(3), 827-882. Retrieved from <http://ssrn.com/abstract=1935617>
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6, 332-339. doi: 10.1038/nrn1651
- Schwartz, J. L. (2013). Evidence and ethics in mandatory vaccination policies. *American Journal of Bioethics*, 13(9), 46-48. doi: 10.1080/15265161.2013.815023

- Sellaro, R., Derks, B., Nitsche, M. A., Hommel, B., van den Wildenberg, W. P., van Dam, K., & Colzato, L. S. (2015). Reducing prejudice through brain stimulation. *Brain Stimulation*, 1-7. doi: 10.1016/j.brs.2015.04.003
- Sergeeva, E. G., Henrich-Noack, P., Bola, M., & Sabel, B. A. (2014). Brain-state-dependent non-invasive brain stimulation and functional priming: A hypothesis. *Frontiers in Human Neuroscience*, 8, 1-2. doi: 10.3389/fnhum.2014.00899
- Siegel, J. Z., & Crockett, M. J. (2013). How serotonin shapes moral judgment and behavior. *Annals of the New York Academy of Sciences*, 1299, 42-51. doi: 10.1111/nyas.12229
- Sommers, S. R. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology*, 90(4), 597-612. doi: 10.1037/0022-3514.90.4.597
- Sommers, S. R., & Ellsworth, P. C. (2001). White juror bias: An investigation of prejudice against Black defendants in the American courtroom. *Psychology, Public Policy, and Law*, 7(1), 201-229. doi: 10.1037/1076-8971.7.1.201
- Special Committee on Standards of Judicial Conduct of the American Bar Association. (2014, March 20). *Code of conduct for United States judges*. Retrieved from <http://www.uscourts.gov/judges-judgeships/code-conduct-united-states-judges>
- Stanton-Ife, J. (2009). The limits of law. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2009 ed.). Retrieved from <http://plato.stanford.edu/archives/spr2009/entries/law-limits/>
- Taylor v. Louisiana, 419 U.S. 522 (1975).
- Terbeck, S., Kahane, G., McTavish, S., McCutcheon, R., Hewstone, M., Savulescu, J., . . . Norbury, R. (2015). Beta-adrenoceptor blockade modulates fusiform gyrus activity to Black versus White faces. *Psychopharmacology*. doi: 10.1007/s00213-015-3929-7
- Tilburt, J. C., Mueller, P. S., Ottenberg, A. L., Poland, G. A., & Koenig, B. A. (2008). Facing the challenges of influenza in healthcare settings: The ethical rationale for mandatory seasonal influenza vaccination and its implications for future pandemics. *Vaccine*, 26, D27-D30. doi: 10.1016/j.vaccine.2008.07.068
- Turner, R. N., & Crisp, R. J. (2010). Imagining intergroup contact reduces implicit prejudice. *British Journal of Social Psychology*, 49(1), 129-142. doi: 10.1348/014466609X419901

U.S. Const. amend. V

U.S. Const. amend. VI

United Nations General Assembly. (1948). *The universal declaration of human rights*. Retrieved from <http://www.un.org/en/documents/udhr/>

Vidmar, N., & Hans, V. (2007). *American juries: The verdict*. Amherst, MA: Prometheus Books.

Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research*, 34(3), 229-243. doi: 10.2190/FLHV-K4WA-WPVQ-H0YM

Ward, J. (2012). *The student's guide to social neuroscience*. New York, NY: Psychology Press.

Wertheimer, A. (1987). *Coercion*. Princeton, NJ: Princeton University Press.

Wertheimer, A. (2002). Liberty, coercion, and the limits of the state. In R. L. Simon (Ed.), *The blackwell guide to social and political philosophy*. Malden, MA: Blackwell.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, 81(5), 815-827. doi: 10.1037/0022-3514.81.5.815

Xu, K., Nosek, B. A., & Greenwald, A. G. (2014). *Race IAT 2002-2013* [Data file]. Retrieved from <https://osf.io/52qxl/>