

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/354687446>

# I Implicit Biases in Trial Settings

Chapter · September 2021

DOI: 10.1007/978-1-4614-7883-6\_668-2

CITATIONS

0

READS

264

2 authors:



Goran Dominioni

Dublin City University

56 PUBLICATIONS 573 CITATIONS

SEE PROFILE



Evangelia Nissioti

4 PUBLICATIONS 2 CITATIONS

SEE PROFILE

---

## Implicit Biases in Trial Settings



Goran Dominioni<sup>1</sup> and Evangelia Nissioti<sup>2</sup>

<sup>1</sup>Dublin City University, School of Law and Government, Dublin, Ireland

<sup>2</sup>Institute of Law and Economics, University of Hamburg, Hamburg, Germany

### Definition

Implicit racial biases are shifts in judgment caused by automatic and/or unconscious attitudes/stereotypes held towards a racial group. In this context, “automatic” means that the bias occurs without any need for attention and that it is difficult to control, whereas “unconscious” means that introspection does not reveal the attitude/stereotype.

### Introduction

In the last decades, studies in behavioral economics and psychology have highlighted the existence of various behavioral patterns that are inconsistent with the rational choice theory (Kahneman 2011 and references therein). Many of these behavioral patterns are relevant for the study of law and policy-making (Jolls et al. 1998). Part of this literature focuses also on implicit biases (e.g., Jolls and Sunstein 2006; McAdams and Ulen 2009; Teichman and Zamir 2014, Dominioni 2020). Although implicit biases do not relate

only to race (but, for instance, also to gender), most of these studies within behavioral law and economics focus on this particular issue. For this reason, the scope of this entry is restricted to implicit racial biases.

### Scientific Basis and Research Methods

Two core concepts in the study of implicit biases are attitudes and stereotypes. An attitude is a mental association between a racial group and an evaluative disposition (Greenwald and Krieger 2006). This evaluative disposition can be either positive or negative. For example, a person may hold either a positive or a negative attitude towards Asians. Instead, racial stereotypes are an association between a racial group and a positive or a negative characteristic (e.g., being lazy, good in math, or aggressive).

Research in social cognition has studied the relationship between attitudes and stereotypes both at the implicit and the explicit level. This literature indicates that a person can hold an implicit attitude and a stereotype that point in opposite directions. For instance, it is possible to have a negative attitude towards Blacks and yet hold a positive stereotype towards them (e.g., Blacks are good in sports). In addition, existing evidence indicates that self-reported attitudes and stereotypes do not necessarily mirror implicit measures. As a consequence, it is possible for an individual to hold an implicit attitude or

stereotype that conflicts with the ones she consciously endorses (Greenwald and Krieger 2006).

Implicit attitudes and stereotypes are identified using implicit measurement procedures. Thanks to these techniques, it is possible to identify attitudes and stereotypes that, for a variety of reasons, explicit measures or actual behavior may fail to detect (e.g., unawareness) (Forscher et al. 2019). Among the most commonly used techniques to measure implicit biases, there are the implicit association test (IAT), affective priming, and brain imaging.

The most commonly used procedure in the implicit biases literature is the IAT (Greenwald et al. 1998). This procedure allows to test the relative strength of implicit attitudes and stereotypes via the measurement of response latencies in the categorization of stimuli into classes. In particular, in a racial IAT, part of the stimuli relates to racial groups (e.g., a name more commonly associated with a Black/White person), while the remaining stimuli relate to other concepts (e.g., good/bad). Subjects are repeatedly asked to categorize each stimulus (e.g., Black name) as belonging to one of two dyads. Thus, for instance, subjects are first asked to categorize the good/bad concepts and Black/White names as belonging to either the Black/good or the White/bad dyad. And, subsequently, the task is repeated with the Black/bad and White/good dyads. Differences in response time between different combinations of dyads indicate that a certain racial group is more easily associated with a certain concept.

Another widely used measurement procedure is affective priming which allows subjects to be exposed to two types of stimuli – a target and a prime – and they are asked to categorize the target as either positive or negative. To perform the task, the only relevant stimulus is the target; therefore, the prime should not affect the behavior of the subjects. For instance, in studies on implicit racial attitudes, subjects are first exposed to pictures of Black faces and White faces (the prime). This exposure occurs subliminally, meaning that pictures are shown for a very short time (e.g., 200 milliseconds) so that they are only unconsciously processed. Subsequently, subjects are exposed to

a series of words (the target) and are asked to categorize them (positive/negative). When being primed with a racial stimulus facilitates the categorization of the words as negative, the subject is said to hold a negative attitude towards the group (Fazio et al. 1995).

A technique often employed in the implicit racial attitude domain is blood oxygen level-dependent contrast imaging in functional magnetic resonance imaging (Fazio and Olson 2003). This physiological measure identifies variations in oxygen levels in the blood present in different parts of the brain, which are positively correlated with the activation of these areas. Thus, for instance, exposure to Black faces has been shown to generate a greater activation of the amygdala (the amygdala is a part of the brain related to the processing of emotions). Therefore, this strand of research indicates that exposure to different racial groups can trigger distinctive emotional states (Kubota et al. 2012).

## Main Findings

The existing research indicates that implicit racial biases are pervasive in the White population of several Western countries. The largest database regarding the diffusion of implicit biases comes from the online platform Project Implicit, which allows visitors to take a racial IAT online. An analysis of the data gathered on this platform between the years 2000 and 2006 indicates that more than 65% of the people who took the race IAT associated relatively more easily Blacks with bad. In addition, less than 15% of the participants (mainly non-White subjects) held relatively stronger associations between White and bad (Nosek et al. 2007).

Research in implicit social cognition indicates that various factors mediate the formation of implicit biases. In particular, empirical evidence indicates that implicit biases can stem from past (often forgotten) experiences, which may date back to early childhood. These experiences can be either direct or indirect. As indirect experiences – for instance, via the media – generally contribute to the formation of a negative

perception of Blacks, this could explain the absence of a pro-Black bias among Black people (Greenwald and Krieger 2006).

A large part of the literature on implicit biases analyzes their impact on human behavior (both spontaneous and deliberate). Two meta-analyses on racial IAT studies find that measurements of implicit biases have higher predictive validity than self-reported measures (Greenwald et al. 2009; Oswald et al. 2013). A meta-analysis of racial priming studies reaches a similar conclusion (Cameron et al. 2012). A recent meta-study (Forscher et al. 2019) analyzed different procedures that aim to change implicit measures, such as procedures that weaken or strengthen mental associations, invoke specific goals and motivations, or lead to mental depletion. This study finds that these procedures had a measurable impact on implicit measures. Nevertheless, this impact did not mediate variations in explicit measures and behavior (Forscher et al. 2019). Overall, this literature suggests that implicit biases might sometimes affect behavior, but it is unclear under which conditions this effect exists.

## Applications in Trial Settings

Having introduced the background of the topic, let us now analyze how implicit biases relate to trial settings from a law and economics perspective. Broadly speaking, existing research on implicit biases unfolds in two directions. First, studies in implicit social cognition made their ways into law reviews because of the impact that implicit biases have on trial actors' behavior. Second, implicit biases have been widely discussed among legal academics in relation to the epistemic value of implicit measurements in the context of evidence law. The second line of inquiry is *stricto sensu* not behavioral and is thus less relevant for the present discussion. In the following, we will, therefore, focus only on the behavioral side of the debate.

### Implicit Biases and Trial Parties' Decisions

Earlier law and economics literature depicts trial participants (judges, prosecutors, and lawyers) as rational agents that allocate resources following a

rational calculation of the costs and benefits of their actions (McAdams and Ulen 2009). Later developments in the field questioned this simplistic description of trial participants and emphasized the role of institutional details and behavioral aspects of human decision-making. Research on implicit biases is a subset of these developments.

With few exceptions, research on implicit biases in the courtroom has focused on criminal trials (Kang et al. 2012). In these settings, a figure that plays a major role in steering the unfolding of a trial is the prosecutor. Research suggests that implicit racial biases among prosecutors account for at least part of the disparities in incarceration rates between racial minorities and nonracial minorities in the US criminal law system (Smith and Levinson 2011; Kang et al. 2012). And indeed, implicit biases can be particularly effective in distorting human behavior when confronted with decisions that have the following characteristics: (i) allow for some discretion, (ii) have to be taken quickly, and (iii) do not provide accountability mechanisms. The decisions taken by prosecutors often present all these characteristics. A typical example is the decision on whether to charge a suspect and/or what crime to charge. In these cases, implicit stereotypes may influence prosecutors' decisions to the disadvantage of racial minorities. For instance, empirical evidence indicates that many individuals tend to have a stronger implicit association between Blacks and aggressive behaviors than between Whites and aggressive behaviors. Similarly, at the implicit level, Blacks are often more easily associated with guns than Whites. Smith and Levinson (2011) argue that these stereotypes may influence the decision of a prosecutor regarding whether to charge a crime both when the Black person is the alleged victim and when he is the alleged perpetrator. For example, following a shooting, prosecutors may have to decide whether to justify the act for self-defense. When the alleged perpetrator is Black and the victim is White, the bias might make it more plausible in the mind of the prosecutor that the harm caused by the Black was the product of an unjustified aggressive behavior. Instead, when the victim is a Black person, the stereotype may lead the prosecutor to

believe that his aggressive behavior may have justified the actions of the White person under investigation (Smith and Levinson 2011).

Another strand of literature focuses on defense attorneys. In this regard, the concern is twofold. On the one hand, implicit biases may induce a defense attorney to not defend a Black client. This, for instance, may occur when the attorney undervalues the probability of winning a case at trial on the basis of a biased evaluation of the available evidence. On the other hand, implicit biases may impair the performance of an attorney by undermining trust and communication with the client. Contextual factors such as work overload and degree of discretion may enhance the effect of implicit biases on attorneys' decisions. Incidentally, US public defenders often operate with insufficient resources to properly fulfill their tasks (Richardson and Goff 2013). Eisenberg and Johnson provide evidence of the pervasiveness of implicit biases among US defense attorneys (Eisenberg and Johnson 2004).

Most importantly, implicit biases may affect the adjudicators' judgment and decision-making. And indeed, empirical evidence indicates that the judgment of both jury eligible citizens and judges is affected by implicit biases (Hunt 2015; Kang et al. 2012). Regarding judges, Rachlinski and co-authors find that racial IAT scores among 120 US judges show similar patterns to those gathered with other samples. In particular, in line with the results obtained on the Project Implicit platform, they find that the large majority of White subjects held implicit attitudes that favor Whites over Blacks, while Black judges did not show strong preferences towards a particular racial group (Rachlinski et al. 2009). This study also reports that when judges are faced with subliminal primes, implicit biases affect their judicial decisions. However, when they are explicitly framed with the race of the defendant, their implicit biasing is not reflected in their ruling. A similar study conducted with the IAT revealed that judges display implicit biases against Asians and Jews (Levinson et al. 2017).

Implicit biases can influence adjudicators' decisions in many circumstances, and in particular when judges have a relatively higher degree of

discretion. Therefore, the influence of implicit biases might be more pronounced when the evidence presented by the parties is ambiguous (Kang et al. 2012). Hence, it is reasonable to conjecture that the effect of implicit biases might be more severe in civil cases, due to the lower burden of proof (Dominioni 2020). In addition, implicit racial biases can affect memory recalls by leading adjudicators to remember facts in a more stereotypical consistent manner (Levinson 2007). In this regard, a growing strand of literature highlights the existence of various implicit stereotypes that are of particular relevance in a criminal trial context. For instance, various studies suggest the existence of an implicit and bidirectional association between Blacks and crime (Hunt 2015). Similarly, Levinson and co-authors show the existence of a stronger implicit association between Blacks and guilt than Whites and guilt (Levinson et al. 2010). In fact, IAT scores of Black/Guilt association predicted the assessment of ambiguous items of evidence in a mock trial. Along similar lines, in the context of death penalty judgments, implicit associations between racial groups and the value of life have been shown to predict higher rates of death penalty punishments against Black defendants (Levinson et al. 2014). On a related note, implicit biases may partially account for the relatively higher rates of death penalty sentences inflicted on persons of color. For instance, implicit biases may ease finding the behavior of an offender as being heinous, atrocious, or cruel, which is one of the aggravating conditions under which death penalty can be inflicted under US law (Smith and Cohen 2012).

Outside the criminal law sphere, legal scholars have discussed implicit biases in the courtroom mainly with regard to employment discrimination lawsuits (Kang et al. 2012; Gertner and Hart 2012). In this context, the literature mostly focuses on the criteria that US judges are expected to adopt when deciding on the dismissal of a case in its early stage. Here, the main concern regards the *Iqbal* standard, under which a judge should dismiss a case when the intent to discriminate the employee is not the most plausible explanation for the conduct of the employer. In making this evaluation, the standard encourages judges to use their

common sense. Legal scholars argue that, by asking judges to rely on common sense to make this decision, the law facilitates the influence of implicit biases on trial outcomes (Kang et al. 2012; Gertner and Hart 2012). A recent study discusses the effects of implicit racial biases on the functioning of tort trials, highlighting how these biases can lead to over- or under-deterrence in a wide variety of settings and can contribute to such as environmental racism (Dominioni 2020).

### Debiasing and Insulating

Researchers are exploring ways to reduce or eliminate the impacts of implicit biases on trial outcomes. From a general perspective, these interventions can act either on trial actors' general tendencies to be biased or on the environment in which trial decisions are made.

One way to decrease the influence of implicit biases on the general decision-making of trial participants is via training (Rachlinski et al. 2009; Kang et al. 2012; Smith and Levinson 2011; Teichman and Zamir 2014). This training may aim at increasing awareness among trial actors about the existence of implicit biases. In turn, this may have the positive effect of decreasing actors' overconfidence in the objectivity of their own judgment while increasing their motivation to control the biases. Various studies in psychology confirm that motivation to control implicit biases can be effective in reducing their influence on behavior, and hence jurors should be advised to take the perspective of the out-group person (Kang et al. 2012). Rachlinski and co-authors further suggest that judges should take a racial IAT (Rachlinski et al. 2009). Besides increasing awareness among judges, taking a racial IAT test may provide judges with a rough estimation of their biases and may therefore help them take decisions regarding the training they may need as well as avoid overcorrections (i.e., shifting the bias against Whites).

Another strategy to reduce the impact of implicit biases is exposing trial actors to counter-stereotypical information or increasing their contacts with members of racial minorities (Rachlinski et al. 2009; Kang et al. 2012; Smith and Levinson 2011). An obvious path is

increasing racial diversification in the courtroom and in prosecutors' offices, but it can only be effective as a long-term strategy. In the short term, de-biasing can be attempted by introducing images of positive figures of minority members in judges' offices.

Alternatively, interventions could focus on the context in which decisions are made. For instance, trials could be structured so that trial agents are given sufficient time to make their decisions. And indeed, this could reduce their reliance on automatic processes and thus the influence of implicit biases on their decision-making. Similarly, research indicates that conditions of high cognitive load (i.e., situations in which the cognitive activity imposed on working memory is high) ease the influence of implicit biases on judgment (Kang et al. 2012). Therefore, various authors suggest trial actors to adopt strategies that may reduce cognitive load when making decisions (Kang et al. 2012). Additional strategies include hiding racial information from the file that prosecutors use to decide whether to charge for a crime (Smith and Levinson 2011) and increasing jury diversity (Kang et al. 2012; Levinson et al. 2017). In fact, research indicates that racially diverse juries tend to endure more thorough deliberations than all-White juries. In turn, this may reduce the reliance on automatic processes in decision-making.

Accountability can also play a role in reducing the effect of biases. For instance, it has been proposed that prosecution offices could data regarding the racial composition of individuals at each stage of the charging phase to provide useful feedback to prosecutors (Smith and Levinson 2011). Similarly, judges should keep track of their decisions, in order to spot potential systematic biases in their decision-making (Kang et al. 2012; Rachlinski et al. 2009). Another path to increase judges' accountability is increasing the depth of the scrutiny allotted to appellate courts in situations where the racial composition of the court of first instance fuels the suspicion that the decision might be biased (Rachlinski et al. 2009).

A recent addition to the debiasing discussion has been the use of Artificial Intelligence (AI) and machine learning systems that can process large

data sets with the aim of producing better predictions for criminal decision-making. Although the use is not widespread, AI decision systems have been considered useful for parole and bail decisions in criminal trials. On the account that these systems may also generate implicitly biased decisions, debiasing methods such as explainable artificial intelligence have been proposed (Deeks 2019).

Potential interventions to reduce the effect of implicit biases at trial are however not limited to public institutions, as also attorneys can implement strategies to help them overcome their biases. Public defenders could refer to objective standards for triage and checklists to evaluate their cases. These measures may help attorneys to impose clearer limits to their own discretion (Richardson and Goff 2013). Lastly, a recent study argues that the literature on implicit racial biases may suggest relying less on litigation to address negative environmental externalities and using more Pigouvian taxation (Dominioni 2020).

## Future Directions

As discussed in this entry, studies in implicit social cognition can provide many insights on trial dynamics and remedies to contrast racial discrimination in the courtroom. Yet, this field of research is one of the latest developments of the already relatively young field of behavioral law and economics. Therefore, there is still a lot to learn about how these biases operate and how they influence the overall efficiency of the legal system. In the following, we highlight various possible pathways for future research.

First, the literature on implicit biases at trial mainly concentrates on Blacks (and Whites). However, the USA and other countries are becoming prismatic societies, and therefore more attention should be given to other racial and ethnic groups (e.g., Hispanics and Asians) (Levinson et al. 2017). Moreover, European scholarship and policymakers could benefit from research conducted on implicit discrimination against Arabs, Eastern Europeans, and Roma people. Along these lines, it might be warranted to

conduct some studies on implicit stereotypes against Black people in Europe. In fact, as implicit biases are partially a product of culture, it might be interesting to study cross-cultural variations in implicit stereotypes.

Second, another pathway of research that remains largely unexplored regards implicit biases in non-criminal settings (Teichman and Zamir 2014). In particular, there is a need for further research in the fields of civil and administrative law. For instance, it has been argued that implicit biases may affect the application of the standard of proof in civil cases (Hunt 2015). Yet, these hypotheses await testing. On a similar note, it can be useful to examine other dispute resolution fora such as arbitration, mediation, or negotiation. The existence and/or the effect of implicit bias in decisions and agreements reached in Alternative Dispute Resolution is a topic that remains uncovered.

Future research can also contribute to a better understanding of debiasing mechanisms. As shown above, the existing literature has started delving in this direction. Results reached so far are informative, but it is still not clear whether, and to what extent, these interventions are effective in real-life settings (Teichman and Zamir 2014). For instance, as discussed above, one of the main recommendations to decrease implicit biases among judges relates increasing diversity in judicial bodies. Yet, Rachlinski and co-authors found little differences in the pervasiveness of implicit biases between groups of judges coming from jurisdictions with great differences in racial diversity in the judiciary (Rachlinski et al. 2009). Another notable shortcoming of existing research is the untested short-term versus long-term effect of proposed debiasing techniques (Forscher et al. 2019). Based on the above, further research could explore in more depth the conditions under which policies aimed at reducing implicit discrimination at trial are more likely to succeed.

Last, many studies both on implicit biases in general (Forscher et al. 2019) and in the courtroom, in particular, have been conducted with university students. The field would benefit from having research conducted with professionals involved in trials or the general population that



is eligible for jury duty. In particular, on the one hand, it would be interesting to gather data regarding the pervasiveness of implicit biases among trial agents. On the other hand, it would be important to analyze whether expertise affects the degree by which implicit biases impact behavior in trial settings.

## Cross-References

- [Behavioral Law and Economics](#)
- [Bounded Rationality](#)
- [Cognitive Law and Economics](#)
- [Judicial Decision-Making](#)

## References

- Cameron CD, Brown-Iannuzzi JL, Payne BK (2012) Sequential priming measures of implicit social cognition a meta-analysis of associations with behavior and explicit attitudes. *Personal Soc Psychol Rev* 16(4): 330–350
- Deeks A (2019) The judicial demand for explainable artificial intelligence. *Columbia Law Rev* 119(7): 1829–1850
- Dominioni G (2020) Implicit racial biases in tort trials. In: Dominioni G (ed) *Biased trials*. Springer Gabler, Wiesbaden, pp 65–104
- Eisenberg T, Johnson SL (2004) Implicit racial attitudes of death penalty lawyers. *DePaul Law Rev* 53(4): 1539–1556
- Fazio RH, Olson MA (2003) Implicit measures in social cognition research: their meaning and use. *Annu Rev Psychol* 54:297–327
- Fazio RH, Jackson JR, Dunton BC, Williams CJ (1995) Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *J Pers Soc Psychol* 69(6):1013–1027
- Forscher PS, Lai CK, Axt JR, Ebersole CR, Herman M, Devine PG, Nosek BA (2019) A meta-analysis of procedures to change implicit measures. *J Pers Soc Psychol* 117(3):522–559
- Gertner N, Hart M (2012) Employment law: implicit bias in employment discrimination litigation. In: Levinson JD, Smith RJ (eds) *Implicit racial biases across the law*. Cambridge University Press, New York, pp 80–94
- Greenwald AG, Krieger LH (2006) Implicit bias: scientific foundations. *Calif Law Rev* 94(4):945–967
- Greenwald AG, McGhee DE, Schwartz JL (1998) Measuring individual differences in implicit cognition: the implicit association test. *J Pers Soc Psychol* 74(6): 1464–1480
- Greenwald AG, Poehlman TA, Uhlmann EL, Banaji MR (2009) Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *J Pers Soc Psychol* 97(1):17–41
- Hunt JS (2015) Race, ethnicity, and culture in jury decision making. *Annu Rev Law Soc Sci* 11:269–288
- Jolls C, Sunstein CR (2006) The law of implicit bias. *Calif Law Rev* 94(4):969–996
- Jolls C, Sunstein CR, Thaler R (1998) A behavioral approach to law and economics. *Stanford Law Rev* 50(5):1471–1550
- Kahneman D (2011) *Thinking, fast and slow*. Macmillan
- Kang J, Bennett M, Carbado D, Casey P (2012) Implicit bias in the courtroom. *UCLA Law Rev* 59(5): 1124–1187
- Kubota JT, Banaji MR, Phelps EA (2012) The neuroscience of race. *Nat Neurosci* 15(7):940–948
- Levinson JD (2007) Forgotten racial equality: implicit bias, decisionmaking, and misremembering. *Duke Law J* 57(2):345–424
- Levinson JD, Cai H, Young D (2010) Guilty by implicit racial bias: the guilty/not guilty implicit association test. *Ohio State J Crim Law* 8(1):187–208
- Levinson JD, Smith RJ, Young DM (2014) Devaluing death: an empirical study of implicit racial bias on jury-eligible citizens in six death penalty states. *NYU Law Rev* 89(2):513–581
- Levinson JD, Bennett MW, Hioki K (2017) Judging implicit bias: national empirical study of judicial stereotypes. *Florida Law Rev* 69(1):63–114
- McAdams R, Ulen T (2009) Criminal behavioral law and economics. In: Garoupa N (ed) *Criminal law and economics*. Edward Elgar, Cheltenham, pp 403–436
- Nosek BA, Hansen JJ, Devos T, Lindner NM, Ranganath KA, Smith CT, Olson KR, Chugh D, Greenwald AG, Banaji MR (2007) Pervasiveness and correlates of implicit attitudes and stereotypes. *Eur Rev Soc Psychol* 30(1):36–88
- Oswald FL, Mitchell G, Blanton H, Jaccard J, Tetlock PE (2013) Predicting ethnic and racial discrimination: a meta-analysis of IAT criterion studies. *J Pers Soc Psychol* 105(2):171–192
- Rachlinski JJ, Johnson SL, Wistrich AJ, Guthrie C (2009) Does unconscious racial bias affect trial judges? *Notre Dame Law Rev* 84(3):1195–1246
- Richardson LS, Goff PA (2013) Implicit racial bias in public defender triage. *Yale Law J* 122(8):2626–2650
- Smith RJ, Cohen GB (2012) Capital punishment: choosing life or death (implicitly). In: Levinson JD, Smith RJ (eds) *Implicit racial biases across the law*. Cambridge University Press, New York, pp 229–243
- Smith RJ, Levinson JD (2011) Impact of implicit racial bias on the exercise of prosecutorial discretion. *Seattle Univ Law Rev* 35(3):795–826
- Teichman D, Zamir E (2014) Judicial decisionmaking: a behavioral perspective. In: Zamir E, Teichman D (eds) *The Oxford handbook of behavioral economics and the law*. Oxford University Press, New York, pp 664–702