Racial Bias and In-Group Bias in Virtual Reality Courtrooms

Samantha Bielen   *Hasselt University*

Wim Marneffe   *Hasselt University*

Naci Mocan   *Louisiana State University*

Abstract

We filmed videos of criminal trials using three-dimensional virtual reality (VR) technology, prosecuted by actual prosecutors and defended by actual defense attorneys in a real courtroom. This is the first paper that utilizes VR technology in a non-computer-animated setting. We alter only the race of the defendants, holding all activity in the courtroom constant, to create arguably perfect counterfactuals. Law students and economics students made conviction and sentencing decisions in these trials that differed only in defendants' race. White evaluators are harsher toward minority defendants in both conviction and sentencing. Minority evaluators are harsher toward minorities in conviction but more lenient in assigning prison terms. This pattern of behavior leads to significant bias against minorities at all stages—conviction, prison sentence, and fine—which is partly a reflection of the numerical majority of the evaluators being white. The same racial bias is observed in the decisions of practicing attorneys.

1. Introduction

The investigation of racial bias in decision-making is important for both scientific inquiry and public policy. A particularly important decision is judgment about others with consequential outcomes. For example, although judicial decisions are expected to be made blindly, whether such impartiality exists in practice has long been debated, and a great deal of research has focused on the in-

vestigation of biases against minorities in conviction, sentencing, and clemency decisions made by judges, juries, and state governors (Alesina and La Ferrara 2014; Rehavi and Starr 2014; Abrams, Bertrand, and Mullainathan 2012; Argys and Mocan 2004; Glaeser and Sacerdote 2003). This research is important because these are high-stakes decisions with economic consequences, and they are made with deliberation. A related and more nuanced line of research recently emerged to investigate the existence of in-group bias (a decision maker's preferential treatment of the members of his or her own group) (see Depew, Eren, and Mocan 2017; Grossman et al. 2016; Anwar, Bayer, and Hjalmarsson 2012; Shayo and Zussman 2011; Gazal-Ayal and Sulitzeanu-Kenan 2010).

The analysis of judicial decisions is complicated by a number of inherent empirical problems related to omitted variables and selection. Although random assignment of defendants to judges alleviates some of the selection problems (Eren and Mocan 2018; Depew, Eren, and Mocan 2017; Abrams, Bertrand, and Mullainathan 2012; Shayo and Zussman 2011), random assignment does not resolve all selection issues. Even if defendants are randomly assigned to judges, there are a number of intervening steps taken by prosecutors and defense attorneys before defendants appear in front of judges, and this process may confound the inference obtained from the analysis of judicial decisions.[1] Similarly, the composition of cases that are adjudicated can also change because of the strategic behavior of defense attorneys.[2] The impression of judges' racial bias may also be a reflection of omitted variables. For example, if prosecutors are more diligent and aggressive toward a certain type of defendant in comparison with others, this differential effort in the courtroom can have a systematic effect on judges' decisions. The effort and effectiveness of defense attorneys may also systematically differ between groups of defendants. To the extent that such effectiveness cannot be measured, it creates an omitted-variable bias.[3] Thus, racial differences in judicial decisions

[1] After cases are (randomly) assigned to judges, prosecutors enter the process, and they have the discretion to determine the charges levied against the defendants. The attitude of the prosecutor regarding charging a particular defendant more or less aggressively has an impact on the eventual outcome of the case.

[2] For example, after the case is assigned to a judge, the defense attorney may be more or less open to plea bargaining on the basis of the judge's attributes, such as race and reputation.

[3] For instance, minority defendants may not be able to afford high-quality attorneys, and ineffective defense provided by low-quality attorneys translates into worse outcomes for minorities such as higher probabilities of conviction and longer sentences (Anderson and Heaton 2012). Part of the negative outcomes associated with courts assigning attorneys to low-income defendants is attributable to the attorneys' effort, motivated by their compensation structure (Agan, Freedman, and Owens 2021).

could be driven by the actions of prosecutors and attorneys, not by the racial bias of judges.[4] It is nearly impossible to control for pretrial decisions and for unobservable (to the researcher) actions of various parties in the courtroom (ranging from the style of language used by the defense and the prosecution to the body language of the defendant).

In this paper we eliminate biases caused by these confounders. Using virtual reality (VR) technology, we record trials in a real courtroom with actual prosecutors and defense attorneys. This is the first paper that uses three-dimensional (3D) VR technology with real people as opposed to computer-animated scenes. The VR technology allows us to replace the defendants in the 3D VR videos of the trials, holding constant every spoken word and every action in the court, which enables us to create arguably perfect counterfactuals to identify the effect of a defendant's race. A total of 201 evaluators made decisions about conviction, prison sentence, and fine in accordance with the guidelines provided by the relevant penal code. To guard against sample selection bias in the sentencing phase, we implement a trimming procedure to remove the marginal convicted defendants, who were likely discriminated against in the conviction phase, from the analysis of sentencing. The evaluators are master's-level law students, master's-level economics students, undergraduate economics students, practicing criminal attorneys, prosecutors, and judges. By design, the race of the defendant is uncorrelated with the characteristics of both the prosecutor and the defense attorney and with all courtroom activity. The defendant's race is also uncorrelated with the evaluators' attributes.

We hold the prosecutor and the defense attorney and their actions in the courtroom constant. We do not reveal the names of the defendants to evaluators. The defendants speak only three times during the trial to answer the questions posed by the judge, as they would typically do in such trials. These answers are "yes," "not guilty," and "no." All actor-defendants were born in Belgium and speak Dutch fluently with no accent. This implies that foreignness or the minority status of the defendants can be inferred only from their darker skin complexion and not from their names or accents. As explained in Section OA6 of the Online Appendix, a separate group of 89 students, who were not involved in the experiment, confirmed that the defendants are easily identified as being white or minority from pictures of their faces.

We find that verdicts about the defendants' guilt versus innocence and decisions about sentence severity are not race blind. There is statistically and economically significant overall racial bias against minorities at all stages: conviction, prison sentence, and fine. We also uncover behavioral differences between white and minority evaluators. White evaluators reveal positive in-group bias during all stages of decision-making; that is, they treat white defendants more favorably than minority defendants during the conviction decision and when as-

---

[4] In addition, each case brought to trial has its own set of mitigating and aggravating circumstances, features of the background and criminal history of the defendant, and circumstances of the victim. Not being able to adjust for these factors constitutes another omitted-variable problem.

signing a prison sentence and fine. A more nuanced picture emerges in the case of minority evaluators, who treat minority defendants more harshly during the conviction decision but more leniently during sentencing. Thus, minority defendants are treated more harshly by both white and minority evaluators during conviction. In contrast, defendants receive favorable treatment from in-group evaluators during sentencing. When we investigate whether disparate treatment of defendants is impacted by evaluators' concerns about terrorism, we find no evidence that such concerns significantly impact racial biases in these decisions.

The analysis of attorneys' decisions reveals the same pattern of racial bias we find in the sample of economics students and law students. We also have a sample of judges and prosecutors[5] who made conviction and sentencing decisions on these cases. Although the decisions of judges and prosecutors cannot be analyzed separately because of the small sample size, when we add judges and prosecutors to the sample of attorneys, we obtain similar results as those obtained from the sample of attorneys. The samples of students and attorneys have different strengths. The sample of students is larger and allows us to investigate in-group bias. The sample of attorneys is smaller, and in-group bias cannot be analyzed because all attorneys are white. Attorneys' decisions, however, have more external validity.

2. The General Idea and the Contribution of the Paper

Imagine that we are watching a trial. The district attorney presents his case the way he deems appropriate. The defense attorney defends her client with the knowledge and the skill she possesses. The facts of the case are presented and debated by both parties. The defendant, a white man who sits in front of the judge, listens to the arguments and counterarguments made by the prosecutor and by his defense attorney. At the end of the trial, the judge renders a verdict of conviction or acquittal, and if she finds the defendant guilty, she assigns a prison sentence and/or a fine.

Now imagine that we go back in time and watch the same trial: the same prosecutor, the same defense attorney, the very same arguments, the same words, and the same body language. Everything that took place in the courtroom is precisely the same with one difference: the defendant is someone else. He is not a white man; he has dark skin. Would the judge's decision about the verdict, the prison sentence length, and the fine be different? Are these judicial decisions influenced by whether the judge and the defendant have similar racial backgrounds? If so, are these influences stronger for white judges or minority judges? If there are race-driven effects on these decisions, are they impacted by whether the evaluator believes that terrorism is a major problem in the country? These are the questions we try to answer in this paper.

Of course, if the judge were to watch the same trial the second time (the only difference between the two versions being the defendant's race), she would im-

---

[5] The group of prosecutors does not include those who participated in the filming of the VR trials.

mediately recognize that this is the same case she adjudicated before. Therefore, while the first evaluator watches the trial with the white defendant, we ask another evaluator to watch the same trial with the minority defendant. Each of the six criminal trials analyzed has a different defendant who faces a different burglary or assault charge with different circumstances. Each of the six trials has two versions. The only difference between the two versions is that the defendant in version 1 of each trial is a young white man and the defendant in version 2 of the same trial is a young man with darker skin. We randomly assign these trials to 201 evaluators, making sure that everyone watches all six trials in random order and that everyone watches three trials with white defendants and three other trials with minority defendants. Half of the evaluators watch version 1, while the other half watch version 2 of each VR video.

The VR technology allows us to video record the courtroom activity in 3D. It also allows us to replace the defendant in a VR video with another individual who is video recorded separately. In the end, this production generates two VR videos that are identical except for one detail: version 1 contains the white defendant, and version 2 contains the minority defendant.

Figure OA1 in the Online Appendix displays a comparison of a scene from versions 1 and 2 of videos one of the trials. The version in the top frame involves a white defendant, and the version in the bottom frame involves a minority defendant. In each frame, the person on the right is the prosecutor presenting his case. The defendant is sitting in the middle, and the person on the left, behind the defendant, is the defense attorney. This picture shows how the judge (the evaluator) observes the trial, although watching the VR videos using 3D headsets produces a much more realistic sense of the actions and sounds of the courtroom.

Two short segments of the trial shown in Figure OA1 present a 40-second segment of the fourth trial in two versions on a split screen (see Online Video OV1). The top segment of the screen displays the trial with the white defendant, and the bottom segment displays the same trial with the minority defendant. Every detail in both videos is identical with the exception of the defendants. Online Video OV2 presents the full version of the same trial with the minority defendant, and Online Video OV3 displays the same trial with the white defendant.

Online Video OV4 provides a glimpse of the first trial with the white defendant (note that the prosecutor, the defense attorney, and the defendant are different). Two short segments are included in the clip. The first segment is from the prosecutor's opening remarks, and the defense attorney speaks in the second part. The prosecutors and the defense attorneys in all videos are prosecutors and defense attorneys who practice criminal law in the province of Limburg in Belgium. We used a real courtroom in Hasselt. The videos are filmed from the judge's bench so that the evaluators could observe the trial from the judge's vantage point. At the conclusion of each trial, each evaluator rendered a verdict of guilt or innocence and assigned a sentence if he or she found the defendant guilty. We compare the conviction decisions and the prison sentence lengths and fines between the evaluators who watched the same trial but faced a white defendant versus a mi-

nority defendant. This design enables us to identify the impact of the race of the defendant on the decisions, holding all else constant. Furthermore, because we have information about the evaluators' race, we investigate whether evaluators are more lenient toward defendants who are part of their in-group.[6]

Previous experimental research that investigates the impact of defendants' characteristics (for example, race or mental illness) on trial verdicts provided mock trial transcripts to be read by the evaluators. In those studies, the race of the victim or defendant was manipulated using photographs and names (Mossière, Maeder, and Pica 2018; Maeder, Mossière, and Cheung 2012), or evaluators were given vignettes that describe the offender and the circumstances of the case (Mercado, Bornstein, and Schopp 2006). In some studies, participants were shown videos of mock trials in which the trial conditions, such as eyewitness evidence, were manipulated by the experimenters (Jones et al. 2017).

It is also informative to compare our design with that of audit studies in which pairs of individuals (auditors) with similar or identical characteristics with the exception of one attribute (for example, race, gender, or criminal record) are sent to apply for a job (Pager, Western, and Sugie 2009; Pager 2003; Neumark, Bank, and Van Nort 1996), to buy a car (Ayres and Siegelman 1995), to buy a housing unit (Ondrich, Ross, and Yinger 2000), and so on. These matched auditors are trained to respond to and interact with the decision maker in the same manner. The difference in outcomes between the matched pairs is attributed to decision makers' discriminatory behavior related to the attribute that differs between the pairs. A primary weakness of audit studies is the difficulty in matching the auditors so that they are identical in all characteristics that might matter for the decision maker.[7] Importantly, auditors are typically aware of the purpose of the study, and as highlighted by Riach and Rich (2002), Heckman (1998), and Turner, Fix, and Struyk (1991), they may consciously or unconsciously behave and interact with the decision maker in a manner that may generate evidence for discrimination. These concerns are alleviated in our study because, unlike typical audit studies in which the actors interact and have conversations with the decision makers to buy or rent something or to interview for a job, our actor-defendants do not interact with the evaluators. This is organic and natural in our setting, because the defendant's representation is provided by his attorney, who does nearly all the talking in the courtroom: our actor-defendants speak only three times to say "yes," "not guilty," and "no." Also, unlike audit studies, everything that should matter for the outcome is identical between the white and minority defendants, including the details of the case and the behavior of the prosecutor and the defense attorney.[8]

---

[6] This component could not be implemented in the sample of attorneys because all attorneys were white. The same was true for judges and prosecutors who evaluated the cases.

[7] As described by Heckman (1998) and Heckman and Siegelman (1993), auditors may also differ on dimensions that are unobservable to the designer of the audit study but that are at least somewhat visible to the decision maker and relevant for the decision. These unobservable characteristics and their distribution between groups can produce differences in outcomes.

[8] A related strand of literature is based on correspondence studies in which pairs of fictitious résumés, letters of interest, or applications are sent to prospective employers (or other decision makers) and the documents of the pairs differ in one attribute, such as the name (signaling race), age,

There exists research that aims to differentiate between theories and mechanisms of racial bias. For example, List (2004) reports an audit study with buyers and sellers of collectible baseball cards and then combines it with dictator games conducted in the lab with the card dealers. Levitt (2004) analyzes the behavior of contestants on the game show *Weakest Link,* and Bertrand and Mullainathan (2004) analyze in a correspondence study the extent to which callbacks of job applicants are responsive to the information about race contained in their names and the amount of information provided about their qualifications and credentials. In a recent empirical study, Arnold, Dobbie, and Hull (2020) analyze judges' bail decisions in New York City. Their data set contains extraordinary detail on defendants, including criminal history and past pretrial misconduct (failing to appear in court or being arrested for a new crime after having been released on bail) and the conduct of the defendant after bail has been granted by the judge. Leveraging the detail in the data and imposing a structure on the manner in which judicial decisions are made, Arnold, Dobbie, and Hull (2020) shed light on the theoretical channels driving the racial discrepancy in bail decisions.

Our paper does not attempt to tease out the extent to which the bias is driven by taste-based or statistical discrimination because our research design does not permit a meaningful analysis along these dimensions. Nevertheless, in an effort to provide more insight into the source of the identified bias, we expand the analysis. For example, it could be argued that racial biases identified in the paper might be driven by certain characteristics of individual defendants that are visible to evaluators, such as defendants' perceived trustworthiness or their perceived proclivity for delinquency. We show, however, that accounting for perceived criminal proclivity or perceived honesty of defendants (assessed by a different group of students who are not involved in the experiment) does not alter the results.

Nine to 15 days after the experiment, the evaluators completed an online survey of 40 questions about a variety of subjects, including whether they consider terrorism an important issue. The results of our paper are similar for those who think that it is very important that the Belgian society pays attention to terrorism and those who were less concerned.

3. The Design of the Study and the Data

We were granted permission by the district attorney of Limburg to use actual criminal cases, prosecutors, and a courtroom. After discussions with prosecutors, we decided to focus on burglary and assault cases because such cases provide substantial discretion to the judge regarding sentencing, and no special background or additional expert information is needed to deliver a verdict. Three burglary and three assault cases were selected from the court archives or the docket.

Two prosecutors who specialize in burglary and assault cases volunteered to participate. We contacted a number of law firms specializing in criminal cases

---

or sex. Differences in the callback rates are attributed to the difference in the attributes. A detailed discussion of correspondence studies is provided by Bertrand and Duflo (2017).

to be part of the experiment. After interviews, we selected three lawyers (two females and one male) with experience in both burglary and assault cases. The lawyers and the prosecutors were given the case files 1 week before the filming of the
videos and were asked to prepare the cases as they would for an actual court trial.
The lawyers did not talk to the prosecutors before the filming to keep the trial
environment as realistic as possible. The prosecutors and the lawyers tried and
defended the cases in the way they would normally in real-life trials. There was no
script to follow, and they presented their cases with no instructions from anyone.

In Belgium the judge is in control of the court hearing, although his or her role
is limited. The judge talks very little during the trial and speaks only to direct
the proceedings. The judge first asks the defendant if he understands the charges
against him and whether he pleads guilty or not guilty. In each of our videos,
the defendants indicate that they understand the charges and that they plead not
guilty. The judge then allows the prosecutor to start with his statement. Once the
prosecutor is finished, the defense attorney launches his or her pleading.[9]

For the sake of realism in the videos, we needed a judge to control the flow of
proceedings. We selected a retired judge from the region of Hasselt to control
the cases as he would normally. The VR camera was located in front of his bench.
Thus, the evaluators who watched the videos could hear his voice in the VR video
but could not see him. This way the setting remained realistic, and the evaluators
viewed the cases from the perspective of the judge. We used one of the courtrooms in the main courthouse in Hasselt.

We employed three white defendants and three nonwhite defendants for our
six trials.[10] We limited the number of trials to six because on average a trial took
12 minutes, which meant that watching the six VR videos and making decisions
for the six trials would take 2 hours. Asking the evaluators to watch trials for more
than 2 hours was infeasible for a number of reasons. There was a 5-minute break
after watching the first three videos and making judgments on these cases. Interactions with other participants were not allowed. The evaluators were not primed
in any way. Figure OA2 in the Online Appendix presents pictures of the actors.
Figures OA5–OA10 present additional snapshots of them during the trials.

A total of 153 students from the faculty of business economics and the faculty
of law at Hasselt University participated as evaluators. The 86 economics students
were a mixture of undergraduate and master's degree students enrolled in a policy evaluation course. We also recruited practicing attorneys from the Limburg

---

[9] In a real hearing, the judge sets the time frame for his or her verdict (normally 4 weeks after the
conclusion of the hearing) and can ask the defendant some questions during the trial. Our setup
(with the evaluators watching the VR videos with headsets) does not allow the evaluators to interrogate the defendant. It is, however, not uncommon in Belgian criminal proceedings for the judge to
ask the defendant no questions.

[10] We selected male actors who live outside of the region where the evaluators reside to avoid
the possibility of an evaluator recognizing one of the actors. We also made sure that all actors were
about the same age, and we gave them instructions on wardrobe so that the clothing worn by the
pairs of actors was similar.

Bar Association, and 36 attorneys decided to participate. Additional information is provided in Section OA2.2 of the Online Appendix.

The students participated in the experiment on November 22–23, 2017. The attorneys participated during a 10-day period in July 2018. There were no unusual news announcements or events in Belgium during this period related to immigration, terrorism, or asylum seekers. It should be remembered, however, that the terrorist attacks in Brussels that targeted the airport and the subway system took place on March 22, 2016, which is 20 (28) months before the students (attorneys) evaluated the cases in this experiment. In addition, the terrorist attacks in Paris, in which most of the perpetrators were residents of Brussels, took place on November 13, 2015. Thus, it is not unreasonable to assume that these major events were remembered by the evaluators.

The evaluators were granted anonymity during the experiment and in the follow-up survey, which they completed online 9–15 days after the experiment (see Section OA5 of the Online Appendix for details). After reading the case folder, evaluators could click on a button on the screen to start the VR video. They had 30 seconds to put on the headsets and watch the video. After watching the video, they took off the headsets and clicked on a button on the screen to go to the deliberation phase (see Figure OA3 in the Online Appendix). First, they were asked to decide on conviction or acquittal. If they decided to convict, they had to assign the sentence (prison sentence and/or fine). As is the case with real judicial decisions, they also had to make a decision about whether and how much to suspend the prison sentence and the fine. These decisions were indicated by clicking and typing on the screen. Details are provided in Section OA4.2 of the Online Appendix.

After watching and rendering decisions for all six cases, participants were presented with an overview of their decisions, and they could alter any decision they made (conviction, prison term, and fine) for any of the cases before they were finished with the experiment. In the analyses, we use the final decisions made for each case.[11] Nine days after the experiment, the evaluators were asked to complete the online survey, which we use to elicit information about their cultural background and attitudes toward social and economic issues.

If the defendants in the VR videos are not recognizable as being a member of a minority or majority group in Belgium, then our design will provide null results.[12] Put differently, if the evaluators could not correctly infer the racial background of our defendant-actors by looking at them, then they would not assign differential verdicts or sentences even if they had racial biases.[13] To analyze whether the

---

[11] To investigate if the evaluators took their task seriously, we conducted a number of tests such as analysis of the speed of the decisions, the order in which the decisions were made, and so on. These are detailed in Section OA9 of the Online Appendix.

[12] Section OA7 of the Online Appendix provides a discussion of race versus ethnicity.

[13] To put it bluntly, the question is whether the evaluators can tell the difference between a Belgian person of European heritage and another individual who is a racial or ethnic minority with darker skin (for example, a Belgian with Moroccan or Turkish origin). Or would the evaluators think that the latter person is a white European who happens to have a suntan?

defendants can be identified as white or minority, we showed pictures of the defendants to a different group of 89 students. As detailed in the Online Appendix (Section OA6), the students identified the defendants' race correctly with 99 percent accuracy.

Because we have six cases (three burglaries and three assaults) and two versions of each case (one with a white defendant and one with a nonwhite defendant) we have 12 videos in the experiment, and each defendant-actor participated in two videos. We made six sets with different sequences of the videos and three different combinations to make sure that each evaluator would see each defendant only once. The sets are described in Online Appendix Table A1.

4. Descriptive Statistics

Table 1 presents the descriptive statistics related to student-evaluators' attributes. Information about evaluators' sex and whether they are law students or economics students was obtained during the experiment. Using the online survey that was completed by students 9 days after the experiment, we measured evaluators' beliefs about a number of social issues and elicited additional personal background information. The variable Minority Evaluator equals one if either the mother or the father of the evaluator was born outside Belgium and if that country is outside the European Union (EU).[14] Thus, minority evaluators are first- or second-generation immigrants whose ancestral origins are in Turkey (10 evaluators), Morocco (four evaluators), Iraq, Zimbabwe, or Armenia (one evaluator each). Alternatively, we determined if the evaluator was a minority in Belgium on the basis of answers to the following question: "Do you or anyone in your household speak a language other than Dutch?"[15] In this alternative definition, we coded the evaluator as a minority if he or she indicated that Arabic, Armenian, Turkish, or Shona is spoken at home.[16] The dummy variable Terrorism Is a Very Important Problem equals one if the individual assigned a value of 7 to the question "On a score of 1–7, to what extent do you think it is important that Belgian society pays attention to terrorism?"

Table 2 presents the descriptive statistics of the decisions made by evaluators. When minority defendants are matched with white evaluators, the conviction rate is higher than when white defendants face white evaluators. Interestingly, the same relatively unfavorable outcome for minority defendants is observed even when they are matched with minority evaluators. Minority evaluators convict minority defendants at a higher rate in comparison with white defendants.

---

[14] If the parent was born in a country that is part of the European Union, such as Germany or France, the evaluator is not considered to be a minority in Belgium. There was one evaluator with parent(s) from Japan, one evaluator with parent(s) from South Korea, and another with parent(s) from India. These three individuals of Asian heritage are excluded from the analysis, but considering them as minorities does not alter the results.

[15] Hasselt University is located in Flanders, the Dutch-speaking part of Belgium.

[16] We do not consider an evaluator to be a minority in Belgium when, for example, French, Polish, or Italian is spoken at home.

Table 1
Descriptive Statistics of Evaluators

	Mean
Minority Evaluator	.11
	(.32)
Male Evaluator	.47
	(.50)
Law Student	.44
	(.50)
Terrorism Is a Very Important Problem	.48
	(.50)
N	153

Note. Standard deviations are in parentheses.

Table 2 also displays the average prison sentence imposed on defendants who are found guilty. Part of the prison sentence can be suspended at the discretion of the evaluator, and Table 2 also presents the average suspended prison sentence among various groups. The difference between the initial prison sentence and the suspended sentence is the effective prison term faced by the convicted defendants. For example, for the sample of all evaluators, convicted minority defendants receive an average of 10.15 months of prison time, and 5.81 months of the initial sentence is suspended. Thus, the effective prison term among minority defendants in this sample is 4.34 months on average.

Table 2 shows that minority evaluators assign prison sentences to minority defendants that are about 2.68 months shorter than those they assign to white defendants. This indicates that minority evaluators treat minority defendants more favorably in assigning a prison sentence, although minority evaluators are harsher toward minority defendants when making a decision of guilt or innocence. White evaluators also exhibit positive in-group bias and assign white defendants prison terms that are about 1 month shorter (3.61 months versus 4.54 months) than those they assign minority defendants. Thus, evaluators provide preferential treatment in assigning prison sentences to in-group defendants, which signifies positive in-group bias in effective prison term.

Table 2 also displays information on fines. There are racial differences in fines by race matching between the defendant and the evaluator, although the differences are not statistically significant. We discuss the details in the empirical analysis.

For the sample of all evaluators, Table 2 shows that minority defendants receive longer sentences than their white counterparts. The same is true for fines, although the differences are not different from 0. This statistical insignificance is misleading because racial bias in conviction decisions generates a selected sample of defendants who are sentenced. Note in Table 2 that while 361 minority defendants are convicted and sentenced, only 309 white defendants are sentenced. These 52 excess minority defendants end up in the sentencing phase because of the 12-percentage-point difference in conviction rates between the races (.79 ver-

Table 2
Descriptive Statistics of Evaluators' Decisions

	All Evaluators		White Evaluator		Minority Evaluator	
	Minority Defendant (1)	White Defendant (2)	Minority Defendant (3)	White Defendant (4)	Minority Defendant (5)	White Defendant (6)
Conviction Rate	.79**	.67**	.78**	.68**	.82*	.65*
	(.41)	(.47)	(.41)	(.47)	(.39)	(.48)
N	459	459	408	408	51	51
Initial Prison Sentence (months)	10.15	9.55	10.47	9.51	7.69	9.82
	(8.62)	(7.57)	(8.85)	(7.42)	(6.15)	(8.91)
N	361	309	319	276	42	33
Suspended Prison Sentence (months)	5.81	5.73	5.93	5.90	4.88	4.33
	(5.20)	(5.57)	(5.26)	(5.62)	(4.72)	(4.94)
N	361	309	319	276	42	33
Effective Prison Sentence (months)	4.34	3.82	4.54+	3.61+	2.81*	5.49*
	(6.73)	(5.11)	(7.03)	(4.93)	(3.48)	(6.31)
N	361	309	319	276	42	33
Initial Fine (€)	782.62	696.70	776.63	685.74	828.10	788.36
	(1,192.96)	(1,088.64)	(1,218.33)	(1,101.48)	(990.45)	(985.24)
N	361	309	319	276	42	33
Suspended Fine (€)	290.87	278.10	288.28	272.96	310.55	321.03
	(559.83)	(519.81)	(557.29)	(528.46)	(585.33)	(445.62)
N	361	309	319	276	42	33
Effective Fine (€)	491.75	418.61	488.35	412.78	517.55	467.33
	(999.70)	(889.49)	(1,027.55)	(906.93)	(764.96)	(737.55)
N	361	309	319	276	42	33

Note. Values are means, with standard deviations in parentheses.
+ $p < .10$.
* $p < .05$.
** $p < .01$.

sus .67).[17] Thus, we present the empirical results both ways: accounting for this selection bias and ignoring the selection bias.

5. Econometric Analysis

5.1. Overall Racial Bias

To investigate the existence of overall racial bias in conviction and sentencing decisions, we run regressions of the following type:

$$Y_{cj} = \alpha_1 + \beta_1 \text{Minority Defendant}_{cj} + \boldsymbol{X}_j \Phi + v_c + \varepsilon_{cj}, \qquad (1)$$

where Y_{cj} is the outcomes related to case (trial) c evaluated by evaluator j. The first outcome is an indicator to represent whether the defendant is convicted by evaluator j. The second and third outcomes are the prison term and the amount of the fine imposed on the defendant if he is found guilty. The variable Minority Defendant equals one if the defendant in case c (in the VR video of trial c) is a racial minority. Recall that a case is identical across all evaluators who watch the VR video of that case, with the exception of the race of the defendant. Six courtroom hearings ($c = 1, 2, \ldots, 6$) are watched and judged by each evaluator j. Three of these trials involve burglary, and the other three are assault cases. In some specifications we include trial fixed effects (v_c), and in others we control for the type of crime on which the trial was based (assault versus burglary). Vector $\boldsymbol{X}$ includes attributes of the evaluators such as their sex and whether they are law students or economics students. Although all evaluators watch the same trials, the order in which the trials are watched is different by design.[18] The estimated coefficient β_1 identifies whether minority defendants are treated differently than their white counterparts in conviction and sentencing decisions.

5.2. In-Group Bias

To test for the existence of in-group bias, we make use of information obtained from the evaluators regarding their cultural heritage. This information allows us to estimate

$$
\begin{aligned}
Y_{cj} = \alpha_2 &+ \beta_2 \text{Minority Defendant}_c + \gamma_2 \text{Minority Evaluator}_j \\
&+ \delta_2 \text{Minority Defendant}_c \times \text{Minority Evaluator}_j + \boldsymbol{X}_j \Theta + \mu_c + \omega_{cj}.
\end{aligned} \qquad (2)
$$

In equation (2) Minority Evaluator is a dichotomous indicator that equals one if

---

[17] The same point is made by Depew, Eren, and Mocan (2017). Because the sequential decision of conviction and sentencing of the defendants by the same judges generates sample selection in the sentencing phase, that study analyzes the sentencing outcomes of defendants who pled guilty (as opposed to those who were convicted and then sentenced by the same judge).

[18] As Table OA1 in the Online Appendix demonstrates, evaluators who were randomly assigned to set 1 or set 2 watched the six cases in a particular sequence (sequence 1), while those who were randomly assigned to sets 3 or 4 watched the same videos in a different order (sequence 2); sequence 3 is the third sequence in which the videos are watched by evaluators. Vector $\boldsymbol{X}$ contains Sequence 1 and Sequence 2, which are dummy variables that control for the sequence in which the videos are watched.

the evaluator who makes the conviction and sentencing decisions is a minority in Belgium. The coefficient γ_2 captures the difference in outcomes between white defendant and minority evaluator pairs and white defendant and white evaluator pairs, and $\gamma_2 + \delta_2$ represents differential sanctions assigned to minority defendants by minority versus white evaluators. Thus, δ_2 is the difference-in-difference estimate that signifies the differential decisions of minority evaluators versus white evaluators in their treatment of minority defendants over white defendants. Put differently, if δ_2 is different from 0, this reflects in-group bias.

While specification (2) identifies in-group bias, it does not provide insights into the source of the bias. One of the contributions of this paper is its ability to unbundle the in-group bias. Because the race of the evaluators is, by design, uncorrelated with all courtroom attributes and case characteristics, and because it is also uncorrelated with the race of the defendant, we can run regressions (3) and (4), conditional on the evaluator's race, to investigate the source of the in-group bias:

$$Y_{cj} = \alpha_3 + \gamma_3 \text{White Defendant} + X_j\Theta + C_c\Lambda + \pi_c + u_{cj} \qquad (3)$$

in the sample of white evaluators and

$$Y_{cj} = \alpha_4 + \gamma_4 \text{Minority Defendant} + X_j\Pi + C_c\Xi + \theta_c + \tau_{cj} \qquad (4)$$

in the sample of minority evaluators.

The coefficient γ_3 reveals whether white defendants are treated differently than minority defendants by white evaluators. The coefficient γ_4 provides the same information for minority evaluators.[19] Note that in equation (2), which analyzes in-group bias, $\alpha_2 + \beta_2$ represents the impact of white evaluators on minority defendants, and α_2 is the impact of white evaluators on white defendants. Thus, β_2 represents the differential treatment of minority defendants vis-à-vis white defendants by white evaluators. Similarly, $\beta_2 + \delta_2$ is the differential treatment of minority defendants by minority evaluators. If p is the proportion of white evaluators in the sample, then the overall differential treatment of minority defendants by all evaluators is equal to $\beta_2 + (1 - p) \times \delta_2$, which should be equivalent to the estimated coefficient of Minority Defendant in equation (1).

6. Results

6.1. Conviction Decisions

The results, reported in Table 3, are based on decisions for conviction versus acquittal by 153 law students or economics students on each of the six cases they watched for a total sample of 918 decisions. (The analyses of attorney decisions

---

[19] Note that γ_3 in equation (3) represents the differential treatment of white defendants by white evaluators and is equivalent to $-\beta_2$ in equation (2). Similarly, γ_4 in equation (4) is the differential treatment of minority defendants vis-à-vis white defendants by minority evaluators, which is represented by $\beta_2 + \delta_2$ in equation (2). Thus, the in-group bias coefficient δ_2 of equation (2) can be constructed by the sum of γ_4 and γ_3.

Table 3
Overall and In-Group Racial Bias in Conviction Decisions

	Overall Bias			In-Group Bias		
	(1)	(2)	(3)	(4)	(5)	(6)
Minority Defendant	.111**	.120**	.120**	.103**	.110**	.110**
	(.028)	(.025)	(.024)	(.029)	(.026)	(.026)
Minority Evaluator				−.025	−.041	−.039
				(.083)	(.076)	(.076)
Minority Evaluator × Minority Defendant				.070	.089	.085
				(.102)	(.087)	(.084)
Law Student	.016	−.011	−.012	.017	−.011	−.012
	(.028)	(.028)	(.028)	(.029)	(.029)	(.029)
Male Evaluator		−.088**	−.095**		−.088**	−.095**
		(.028)	(.029)		(.029)	(.029)
Early Trial	−.127**	−.101**		−.127**	−.101**	
	(.028)	(.030)		(.028)	(.030)	
Sequence 1			−.073*			−.073*
			(.034)			(.034)
Sequence 2			−.037			−.036
			(.031)			(.031)
Assault Case	−.105**			−.105**		
	(.029)			(.029)		
Trial fixed effects	No	Yes	Yes	No	Yes	Yes

Note. The variable Early Trial is a dummy that equals one if the trial is one of the first three trials watched by that evaluator. The variables Sequence 1 and Sequence 2 are dichotomous indicators of the order in which the videos are watched. Standard errors, in parentheses, are clustered at the evaluator level. $N = 918$.
  * $p < .05$.
  ** $p < .01$.

are reported later in the paper.) Table 3 presents the results of estimating equation (1), which investigates the overall racial bias in conviction decisions, and the results of the investigation of in-group bias based on equation (2). In columns 1 and 4 we control for whether the case was an assault or burglary and whether the decision on the case was made during the first half of the experiment or in the second half (Early Trial). Columns 2 and 5 present the results from a specification that includes trial fixed effects, the sex of the evaluator, and whether the decision was made during the first half of the experiment. Columns 3 and 6 control for the sequence type (the order in which videos are watched) in addition to trial fixed effects and the evaluator's sex.

The results for overall racial bias show that minority defendants are about 12 percentage points more likely to be convicted than white defendants. This indicates that two defendants whose cases are identical in every respect except for their race receive different resolutions. That is, the defendants face the same criminal charge with the same mitigating and aggravating circumstances, and they are charged by the same prosecutor and defended by the same attorney in the same

Table 4
Conviction Regressions Conditional
on the Evaluator's Race

	White Evaluator (1)	Minority Evaluator (2)
White Defendant	−.110**	
	(.026)	
Minority Defendant		.204*
		(.076)
Law Student	−.011	−.007
	(.029)	(.139)
Male Evaluator	−.087**	−.092
	(.029)	(.118)
Early Trial	−.121**	.072
	(.031)	(.071)
N	816	102

Note. The variable Early Trial is a dummy that equals
one if the trial is one of the first three trials watched by
that evaluator. Standard errors, in parentheses, are clus-
tered at the evaluator level. All regressions include trial
fixed effects.
 * $p < .05$.
 ** $p < .01$.

way (using same spoken language and same body language in the courtroom).
Nevertheless, minority defendants are about 16 percent more likely to be con-
victed than white defendants.

In the estimation results of equation (2), the coefficient of the interaction term
Minority Defendant × Minority Evaluator represents the impact of in-group
bias in conviction decisions. The point estimate is about .085, although it is not
statistically significant at conventional levels. This provides suggestive evidence
for negative in-group bias: defendants are about 12 percent more likely than the
mean rate to be convicted if they face an evaluator who is of their race. Estimating
the model by including evaluator fixed effects provides very similar results.

Table 4, which presents the estimation results of equation (3) and (4), sheds
light on the source of the statistically insignificant in-group bias reported in
Table 3. White evaluators are 11 percentage points (15 percent) less likely to con-
vict white defendants, which indicates positive in-group bias. On the other hand,
minority evaluators' conviction decisions indicates negative in-group bias: mi-
nority defendants are 20 percentage points (27 percent) more likely to be con-
victed by minority evaluators. Thus, white evaluators exhibit positive in-group
bias, while minority evaluators display negative in-group bias, and the latter ef-
fect is stronger in magnitude. This indicates that the statistically insignificant in-
group bias reported in Table 3 is the result of opposing in-group biases between
racial groups.

Preferential treatment of in-group members is intuitive and unsurprising, and

there is evidence of such behavior (Chen and Li 2009; Shayo and Zussman 2011). Negative in-group bias, however, is also reported in naturally occurring data. Depew, Eren, and Mocan (2017) find that Louisiana juvenile court judges assign longer sentences to defendants who are of their race. Because judges are elected in Louisiana, they may be trying to avoid creating the impression to the voters of favoring their own race, and thus they may be harsher toward these defendants. This is, of course, not the case in our setting because our evaluators do not have such a concern. Another explanation could be that minority evaluators might consider a defendant of a minority group to be a representative of the group. In that case, the evaluator might be harsher toward the defendant for badly representing the group by merely being put in front of the judge, and this sentiment can translate into convicting the defendant even if the case against him in the court hearing is not very strong. Negative in-group bias that can be attributed to this effect has been detected in lab experiments in which in-group members violate a social norm (Mendoza, Lane, and Amodio 2014; Goette, Huffman, and Meier 2006). Similarly, List and Price (2009) report that minority donors are less likely to contribute to a charity during a door-to-door fundraising drive if the solicitor is also a minority and that the gift is smaller, conditional on the decision to give, which suggests lack of trust.

6.2. Sentencing Decisions: Prison Term and Fine

Defendants who are found guilty are sentenced to a prison term and/or fine. Because there is overall racial bias in the conviction decision (see Table 3), this bias contaminates the sample of defendants who are sentenced. More specifically, the existence of racial bias in convictions implies that some minority defendants who should not have been convicted are nevertheless found guilty because of their race. If these individuals represent borderline cases in the decision to convict or acquit, they may receive lenient punishment in the sentencing phase. In that case, ordinary least squares estimates of racial bias in sentence length and fine would be biased downward. To account for such selection, we follow Hoffman and Oreopoulos (2009) and Angrist, Bettinger, and Kremer (2006) and trim the sample that contains those who are found guilty. This procedure assumes that evaluators use a higher standard to determine guilt in cases with white defendants, which implies that the marginal convicted minority defendant is less guilty than the marginal convicted white defendant. Therefore, we remove excess minority defendants (with the lowest levels of punishment) to obtain groups of equal size by race at the sentencing phase. For comparison purposes, we also report the results using the full sample of convicted defendants.

As displayed in Table 2, white evaluators made 408 decisions regarding minority defendants and 408 decisions regarding white defendants. The white evaluators convicted minority defendants 78 percent of the time, while they convicted white defendants with 68 percent probability. This results in 319 convicted and sentenced minority defendants but only 276 convicted and sentenced white

defendants. Put differently, the 10-percentage-point difference in the conviction rates between white and minority defendants, which is due to racial bias of white evaluators, generated 43 excess minority defendants to be sentenced by white evaluators ($319 - 276 = 43$). As shown in the results for conviction rates of minority evaluators in Table 2, minority evaluators too are biased against minority defendants in their conviction decisions (conviction rates of .82 versus .65). This results in nine excess minority defendants ($42 - 33 = 9$) sentenced by minority judges. The trimming procedure allows us to identify these marginal defendants and drop them from the sample. We analyze the sentence distribution of all 319 minority defendants who are convicted by white evaluators and drop the 43 who received the lightest sentences: there are 43 defendants for whom both the effective prison sentence and the effective fine are 0. We applied the same procedure to drop the nine minority defendants (to reduce the number of sentenced minority defendants to the number of sentenced white defendants, from 42 to 33) who were convicted and then sentenced by minority evaluators: those who are at the far left tail of the sentence distribution.[20]

Table 5 presents the estimation results of both the overall and the in-group bias, where the outcome is the effective prison sentence, or the difference between the initial sentence handed down by the evaluator and the suspended sentence (see Table 2). Table 5 presents the results based on the trimmed sample and the full sample (the entire group of convicted defendants). As expected, in the results for the overall racial bias the coefficient of Minority Defendant is larger when the regressions use the trimmed sample. The results indicate that minority defendants receive prison terms about .7 months longer when we do not adjust for the racial selection bias in conviction, but minority defendants receive sentences of 1.3 months longer (about 32 percent of the mean) if we account for selection.[21] Models with evaluator fixed effects provide the same inference. Table 5 reveals the existence of positive in-group bias in prison sentencing: the estimated coefficient of the interaction term Minority Defendant × Minority Evaluator is negative and significantly different from 0 in every specification. The magnitude of the coefficients indicates that if the defendants are sentenced by an in-group evaluator, they receive prison sentences that are almost 4 months shorter. This is a big impact, as the average sentence length is 4 months.

Table 6 displays the results of the models in which the dependent variable is the logarithm of the fine.[22] The results based on the full sample indicate that minority defendants receive fines that are 53 percent higher than those assigned to

---

[20] Six minority defendants convicted by minority evaluators had no effective prison time and no effective fine assigned to them, and they are dropped. Among the remaining convicted minority defendants, we dropped those who had no effective prison time coupled with the lowest effective fines (which were €100 and €300).

[21] As discussed earlier, the racial bias coefficient in Table 5 (the coefficient of Minority Defendant overall bias) is also recoverable from the two coefficients obtained from the in-group bias regressions in Table 5. For example, using column 6 of Table 5, $1.684 - (3.766 \times .11)$ is equal to 1.269 (where .11 is the proportion of minority evaluators in the sample). The coefficient of Minority Defendant in column 3 is 1.285.

[22] Because some effective fines are 0, we added €1 to effective fines assigned by the evaluators.

Table 5
Overall and In-Group Racial Bias in Prison Sentencing

	Overall Bias			In-Group Bias		
	(1)	(2)	(3)	(4)	(5)	(6)
Trimmed sample ($N = 618$):						
Minority Defendant	1.256*	1.300**	1.285*	1.634**	1.689**	1.684**
	(.514)	(.498)	(.511)	(.556)	(.545)	(.556)
Minority Evaluator				1.914	2.103+	2.080+
				(1.266)	(1.205)	(1.226)
Minority Evaluator ×						
Minority Defendant				−3.544**	−3.671**	−3.766**
				(1.136)	(1.211)	(1.240)
Law Student	−.874	−.387	−.372	−.844	−.336	−.327
	(.654)	(.666)	(.670)	(.663)	(.672)	(.678)
Male Evaluator		1.666*	1.751*		1.696*	1.776*
		(.716)	(.754)		(.713)	(.750)
Early Trial	−1.498**	−1.112+		−1.492**	−1.087+	
	(.457)	(.563)		(.457)	(.564)	
Sequence 1			.232			.227
			(.966)			(.966)
Sequence 2			−.298			−.314
			(.647)			(.648)
Assault Case	−3.188**			−3.199**		
	(.428)			(.424)		
Trial fixed effects	No	Yes	Yes	No	Yes	Yes
Full sample ($N = 670$):						
Minority Defendant	.605	.708	.692	1.011*	1.117*	1.114*
	(.470)	(.451)	(.454)	(.508)	(.493)	(.495)
Minority Evaluator				1.911	2.085+	2.077+
				(1.265)	(1.202)	(1.219)
Minority Evaluator ×						
Minority Defendant				−3.638**	−3.720**	−3.832**
				(1.155)	(1.202)	(1.237)
Law Student	−.848	−.345	−.312	−.846	−.323	−.297
	(.617)	(.630)	(.634)	(.627)	(.636)	(.642)
Male Evaluator		1.607*	1.655*		1.629*	1.674*
		(.680)	(.712)		(.676)	(.707)
Early Trial	−1.402**	−1.089*		−1.385**	−1.049*	
	(.427)	(.521)		(.427)	(.523)	
Sequence 1			−.021			−.008
			(.886)			(.890)
Sequence 2			−.278			−.292
			(.627)			(.625)
Assault Case	−3.193**			−3.204**		
	(.411)			(.408)		
Trial fixed effects	No	Yes	Yes	No	Yes	Yes

Note. The variable Early Trial is a dummy that equals one if the trial is one of the first three trials watched by the evaluator. The variables Sequence 1 and Sequence 2 are dichotomous indicators of the order in which the videos are watched. The variable Effective Prison Sentence is equal to the initial prison sentence minus the suspended sentence. Standard errors, in parentheses, are clustered at the evaluator level.

  + $p < .10$.
  * $p < .05$.
  ** $p < .01$.

<div align="center">

Table 6

Overall and In-Group Racial Bias in Fines

</div>

	Overall Bias			In-Group Bias		
	(1)	(2)	(3)	(4)	(5)	(6)
Trimmed sample ($N = 618$):						
Minority Defendant	1.082**	1.061**	1.071**	1.151**	1.131**	1.135**
	(.198)	(.194)	(.196)	(.211)	(.206)	(.209)
Minority Evaluator				1.138+	1.213*	1.216*
				(.599)	(.577)	(.568)
Minority Evaluator × Minority Defendant				−.655	−.664	−.616
				(.571)	(.585)	(.576)
Law Student	−.746**	−.689*	−.695*	−.674*	−.599*	−.604*
	(.270)	(.281)	(.280)	(.275)	(.288)	(.286)
Male Evaluator		.084	.053		.117	.084
		(.284)	(.279)		(.281)	(.277)
Early Trial	.975**	.627*		.962**	.619*	
	(.239)	(.249)		(.240)	(.252)	
Sequence 1			.017			−.010
			(.334)			(.319)
Sequence 2			.078			.081
			(.309)			(.309)
Assault Case	−.098			−.116		
	(.231)			(.232)		
Trial fixed effects	No	Yes	Yes	No	Yes	Yes
Full sample ($N = 670$):						
Minority Defendant	.461*	.436*	.443*	.550*	.523*	.522*
	(.200)	(.198)	(.199)	(.212)	(.210)	(.211)
Minority Evaluator				1.152+	1.220*	1.237*
				(.600)	(.576)	(.576)
Minority Evaluator × Minority Defendant				−.872	−.854	−.781
				(.606)	(.617)	(.605)
Law Student	−.725*	−.681*	−.699*	−.661*	−.599+	−.612+
	(.292)	(.304)	(.303)	(.299)	(.313)	(.311)
Male Evaluator		.112	.057		.142	.086
		(.305)	(.301)		(.304)	(.300)
Early Trial	.945**	.653**		.934**	.645**	
	(.234)	(.242)		(.235)	(.244)	
Sequence 1			−.230			−.256
			(.353)			(.343)
Sequence 2			.035			.038
			(.343)			(.342)
Assault Case	−.255			−.270		
	(.222)			(.223)		
Trial fixed effects	No	Yes	Yes	No	Yes	Yes

Note. The variable Early Trial is a dummy that equals one if the trial is one of the first three trials watched by that evaluator. The variables Sequence 1 and Sequence 2 are dichotomous indicators of the order in which the videos are watched. The variable Effective Prison Sentence is equal to initial prison sentence minus the suspended sentence. Standard errors, in parentheses, are clustered at the evaluator level.

$+ p < .10.$
$* p < .05.$
$** p < .01.$

white defendants. The results based on the trimmed sample reveal that correcting sample selection (stemming from the bias in the conviction decision) more than doubles the estimated coefficient of the Minority Defendant dummy. Results for in-group bias in Table 6 show that the coefficient of the interaction term (Minority Defendant × Minority Evaluator) is negative in all specifications and that the point estimate is around −.6, which indicates that defendants who are of the same race as the evaluator receive fines that are 54 percent lower, although the effect is not statistically significant.[23,24]

Table 7 presents results that unbundle the in-group bias effect in prison sentences and fines. In the trimmed sample, convicted white defendants receive prison sentences that are about 1.7 months shorter than their minority counterparts if they are judged by a white evaluator. Minority defendants receive sentences that are 1.7 months shorter if the evaluator is also a minority (the effect is significant at the 11 percent level). The results based on the full sample reveal the same inference. Thus, evaluators of both races treat in-group defendants equally favorably in the assignment of prison sentencing. Put differently, convicted defendants receive shorter sentences if they are matched with an in-group evaluator, and this bias is driven equally by white and minority evaluators.

Table 7 also shows that, in the trimmed sample, white evaluators fine white defendants more leniently than minority defendants. White evaluators assign fines to convicted white defendants that are 68 percent lower than those they assign to minority defendants (the coefficient of Minority Defendant is −1.128). On the other hand, the impact of the defendant's race on fine is not significantly different for minority evaluators in the trimmed sample. Results for the full sample display a similar picture. Thus, results for in-group bias in Table 6 and for fines in Table 7 reveal positive in-group bias in the assignment of fines. Convicted defendants face lower fines when the person who fines them is of the same race, and this effect is stronger for white evaluators. Models with evaluator fixed effects provide the same results.

In summary, the results reveal positive in-group bias in conviction and sentencing for white evaluators: they are biased in favor of white defendants (or against minority defendants) at all stages of the decision-making. The decisions of minority evaluators are nuanced. They exhibit negative in-group bias in conviction, which indicates that they are harsher toward defendants of their in-group when deciding guilt or innocence. On the other hand, they treat in-group defendants more leniently when assigning prison sentences and fines, although the latter effect is not statistically significant. This is the first paper that identifies such opposing effects in the same decision sequence. Because the majority of the evaluators in the sample are white, the in-group biases translate into overall racial

[23] The impact is calculated as $\exp\{\beta - .5\text{Var}(\beta)\} - 1$, where β is the estimated coefficient and $\text{Var}(\beta)$ is its variance (Kennedy 1981).

[24] We also used an alternative method of correcting sample selection. We assigned no punishment (0 months of prison term and 0 fines) to all defendants who were found not guilty. We then used all defendants, regardless of whether they were convicted, in the sentencing regressions. The estimated impacts were somewhat smaller but statistically significant.

Table 7

Prison and Fine Regressions Conditional on the Evaluator's Race

	Prison Sentence				Fine			
	Trimmed Sample		Full Sample		Trimmed Sample		Full Sample	
	White Evaluator	Minority Evaluator	White Evaluator	Minority Evaluator	White Evaluator	Minority Evaluator	White Evaluator	Minority Evaluator
White Defendant	−1.689**		−1.118*		−1.128**		−.521*	
	(.546)		(.494)		(.206)		(.210)	
Minority Defendant		−1.729		−2.317*		.468		−.302
		(1.034)		(1.057)		(.531)		(.513)
Law Student	−.084	−3.029*	−.112	−2.395+	−.645*	.154	−.673*	.479
	(.706)	(1.276)	(.672)	(1.197)	(.294)	(1.408)	(.320)	(1.451)
Male Evaluator	1.535*	3.072+	1.482*	2.977+	.055	.790	.104	.724
	(.774)	(1.752)	(.740)	(1.667)	(.302)	(.810)	(.326)	(.854)
Early Trial	−1.068+	−1.056	−1.006+	−1.415	.570*	1.019	.623*	1.011+
	(.621)	(1.101)	(.587)	(.939)	(.270)	(.679)	(.268)	(.544)
N	552	66	595	75	552	66	595	75

Note. The variable Early Trial is a dummy that equals one if the trial is one of the first three trials watched by that evaluator. Standard errors, in parentheses, are clustered at the evaluator level. All regressions include trial fixed effects.

+ $p < .10$.
* $p < .05$.
** $p < .01$.

biases in conviction, prison time, and fines against minorities.[25] A number of extensions and robustness analyses do not alter the conclusions. For example, the results are similar between the first three and last three trials. The results are also similar between fast decisions and slow decisions and between decisions that are and are not altered by evaluators at the end of the experiment. These and other sensitivity analyses are summarized in Section OA9 of the Online Appendix.

6.3. Is It Race or Something Else?

Could these findings be an artifact of some other attribute of the defendants? For example, could it be the case that the body language of the defendants is influenced by their race and that it is the body language, rather than the race of the defendants, that triggers the response of the evaluators? There are a number of answers to this question. First, in our six trials and two versions of each trial (generating 12 versions with six white and six minority defendants), body language is very similar between white and minority defendants (see the Online Videos). This is because after filming a trial with a white defendant, the video was shown to the minority defendant who would replace him so that he could mimic the body language of the white defendant.

The body language of the defendants did not project any disrespect toward the judge or the prosecutor.[26] The only possible difference is the white defendant in trial 1, who occasionally crossed his legs, while other defendants did not. If crossing one's legs while sitting in front of a judge is considered to be disrespectful behavior that should be punished, this would imply that this white defendant received harsher punishment than what was appropriate, and it indicates that our results are potentially an underestimate of the racial bias we detect against minorities.

One can argue that minorities might be more timid and nervous during a trial for cultural reasons, while white defendants might be more self-confident. Self-confidence could signal innocence or trustworthiness, while being timid and nervous could suggest guilt, and this could be why the evaluators were biased against minorities. But minority defendants were told to mimic the body language of the white defendants. Second, recall that minority defendants are more likely to be convicted by both white and minority evaluators. Given that minority evaluators are less likely to fall into the trap of cultural misunderstanding of the body language of their in-group, our findings cannot be attributed to possible differences in body language.

---

[25] Recall that, as described at the end of Section 5, the racial bias coefficient β_1 (the coefficient of Minority Defendant) in equation (1) is equal to $\beta_2 + (1 - p)\delta_2$, where β_2 and δ_2 are the relevant coefficients from the in-group bias regression (2) and p is the proportion of white evaluators.

[26] The spoken language did not differ between the defendants. They did not speak during the trial with the exception of their statements regarding their understanding of the charges, their statement about a not-guilty plea, and their statement about not adding anything to their attorney's defense.

Table 8
Bias by the Evaluator's Concern about Terrorism

	Terrorism Important		Terrorism Not Important	
	(1)	(2)	(3)	(4)
Conviction decisions:				
Minority Defendant	.111**	.100**	.133**	.127**
	(.031)	(.032)	(.038)	(.040)
Minority Evaluator		−.038		−.032
		(.140)		(.084)
Minority Evaluator × Minority Defendant		.123		.045
		(.118)		(.124)
N	438	438	480	480
Prison sentencing:				
Minority Defendant	1.004*	1.368**	1.579+	1.942*
	(.491)	(.491)	(.824)	(.898)
Minority Evaluator		2.330		1.463
		(2.150)		(1.451)
Minority Evaluator × Minority Defendant		−3.874*		−2.926*
		(1.917)		(1.369)
N	310	310	308	308
Fines:				
Minority Defendant	1.003**	1.223**	1.094**	.992**
	(.275)	(.286)	(.287)	(.318)
Minority Evaluator		2.564**		.139
		(.482)		(.789)
Minority Evaluator × Minority Defendant		−2.335**		.790
		(.579)		(.752)
N	310	310	308	308

Note. Columns 1 and 3 focus on overall bias, are based on equation (1), and include trial fixed effects, the sex of the evaluator, and whether the decision was made during the first half of the experiment. Columns 2 and 4 focus on in-group bias and are based on equation (2). Standard errors, in parentheses, are clustered at the evaluator level.

* $p < .05$.
** $p < .01$.

6.4. What Is the Driver of the Bias?

It can be conjectured that white evaluators hold minorities responsible (explicitly or implicitly) for undesirable social phenomena. For example, as shown in Table 1, 48 percent of the evaluators are very concerned about terrorism in Belgium. If evaluators who think that terrorism is a very important problem also believe that minorities are more likely to be associated with terrorism, then these beliefs may translate into evaluators' propensity to treat minority defendants relatively harshly. To shed some light on this mechanism, we use the question that was posed to evaluators about terrorism.[27] We ran the models by dividing the

[27] Recall that evaluators answered this question via an online survey 9–15 days after the experiment and that the terrorism question was part of a set of questions including personal characteristics of the respondent and perceptions of institutions, jobs, income, and human values.

sample into two groups of evaluators: those who think that terrorism is a very im-
portant problem in Belgium and those who are less concerned. Table 8 summa-
rizes the results. Column 1 shows that minorities are 11 percentage points more
likely to be convicted than whites if the evaluators think that terrorism is a very
important problem. But column 3 mirrors this result and shows that the minority
impact is 13.3 percentage points among evaluators who are not as concerned
about terrorism.[28] Thus, there is no compelling evidence that evaluators' concerns
about terrorism are the driver of the differential conviction rates between white
and minority defendants. The estimated coefficients are also similar for the in-
group bias effects on conviction decisions in Table 8 and show the same patterns
of behavior in the two groups of evaluators. Consequently, the results for deci-
sions to convict in Table 8 are similar to those obtained for the full sample in
Table 3.

   Table 8 also summarizes the results pertaining to sentence length and fines.
Once again, the results are similar between evaluators who are very concerned
about terrorism and those who are less concerned. Thus, the results displayed in
Table 8 indicate that the racial bias against minority defendants is unlikely to be
driven by evaluators' concerns about terrorism.[29] Of course, as described above,
the terrorist attacks in Brussels and Paris may have been ingrained in the psyche
of the evaluators regardless of how they revealed their concerns about terrorism
in our survey. As a result, dividing the sample between those who are more or
less concerned about terrorism may not lead to a significant difference between
the two groups. On the other hand, it should be recognized that the trials in the
experiment are not related to acts of terrorism, and therefore judgments about
the merits of these cases should, in principle, not be impacted by anxieties about
terrorism.

   Another concern is that some defendants who happen to be white might look
more trustworthy than other defendants to the evaluators. If people who look
more trustworthy are generally treated more favorably, then the racial bias re-
ported here may be attributable to differences in perceived trustworthiness be-
tween white and minority defendants. We did not ask the evaluators about their
assessment of the defendants' trustworthiness because such an inquiry could
have made the evaluators suspicious about the intent of the experiment. Instead,
we showed the pictures of the defendants to a different group of 49 students, who
were juniors at Hasselt University. These students were asked to rate each defen-
dant on the basis of a photo (see Figure OA2 of the Online Appendix) on a scale
from 1 to 7 (1 = very likely to 7 = very unlikely) regarding his perceived trust-
worthiness and criminal proclivity. The questions were "How likely is it that this
person is involved in a small crime such a shoplifting?" and "How likely is it that

---

[28] The average conviction rate in the former group is .78, while it is .69 in the latter, which indi-
cates that the probability of being convicted for minorities is 19 percent higher than that for white
defendants in the latter group of evaluators, while being a minority increases the probability of con-
viction by 14 percent in the former group.
   [29] Using different cutoffs to classify evaluators into groups that are more or less concerned about
terrorism does not alter the results.

Table 9
Descriptive Statistics of Attorney Sample

	Minority Defendant	White Defendant
Conviction Rate	.69*	.55*
	(.47)	(.50)
N	108	108
Initial Prison Sentence (months)	11.2	11.1
	(6.5)	(7.5)
N	74	59
Suspended Prison Sentence (months)	6.2	6.5
	(4.2)	(4.1)
N	74	59
Effective Prison Sentence (months)	5.0	4.5
	(6.5)	(7.5)
N	74	59
Initial Fine (€)	711	701
	(840)	(779)
N	74	59
Suspended Fine (€)	286	284
	(460)	(364)
N	74	59
Effective Fine (€)	425	417
	(554)	(577)
N	74	59

Note. Values are means, with standard deviations in parentheses.
  * $p < .05$.

this person would return a wallet to its owner after finding it on the street?"[30] Accounting for these perceived honesty scores of the defendants did not impact the results in a systematic way, which indicates that the racial bias we report is unlikely to be driven by perceptions of the defendants' honesty or dishonesty.

7. The Analysis of Attorneys' Decisions

Table 9 displays the summary statistics related to the conviction and sentencing decisions made by 36 attorneys who participated in the experiment. They convict at a lower rate than students do, but, consistent with the behavior of students, attorneys too convict minority defendants at a higher rate (69 percent versus 55 percent). The average effective prison sentence is similar between attorneys and students, and average fine assigned by attorneys is only slightly lower (€420 versus €468).[31]

[30] White defendants, as a group, are perceived as more trustworthy. The average value for the three white defendants for the question about returning the wallet to its owner (being unlikely to shoplift) is 4.49 (4.47), while the average value for the three minority defendants is 4.03 (3.94). There is, however, variation between defendants. For example, minority defendant 2 is perceived to be more trustworthy than white defendant 1 and about as trustworthy as white defendant 3. Minority defendant 1 is deemed to be less likely to commit a crime than white defendants 1 and 3.

[31] These are weighted averages of fines assigned to minority and white defendants shown in Tables 2 and 9.

Table 10
Overall Racial Bias in the Attorney Sample

	Conviction		Prison		Fine	
	(1)	(2)	(3)	(4)	(5)	(6)
Minority Defendant	.139*	.139*	.707	.716	.925*	.927*
	(.060)	(.061)	(.890)	(.862)	(.430)	(.421)
Male Evaluator	−.106+	−.108+	2.138	1.840	.569	.431
	(.061)	(.055)	(1.717)	(1.425)	(.656)	(.644)
Early Trial	−.104		−.298		.167	
	(.063)		(1.379)		(.457)	
Sequence 1		.008		1.241		.648
		(.068)		(2.447)		(.758)
Sequence 2		−.149*		.202		.246
		(.061)		(1.022)		(.897)
N	216	216	118	118	118	118

Note. The variable Early Trial is a dummy that equals one if the trial is one of the first three trials watched by the evaluator. The variables Sequence 1 and Sequence 2 are dichotomous indicators of the order in which the videos are watched. Standard errors, in parentheses, are clustered at the evaluator level. All regressions include trial fixed effects.
+ $p < .10$.
* $p < .05$.

Because there is no racial variation in the attorneys (they are all white), in-group bias in their decisions cannot be analyzed. Instead, we focus on the investigation of overall racial bias. Table 10 presents the results for racial bias in conviction decisions and shows that minority defendants are about 14 percentage points more likely to be found guilty by practicing attorneys, holding constant the attributes of the case and trial environment. It is interesting to note that the size of the racial bias coefficient estimated in the sample of attorneys is almost the same as that estimated for the sample of law students (14.6 percentage points, as shown in Online Appendix Table OA2).

Table 10 also presents regression results that analyze the impact of the defendant's race on the prison sentences and fines assigned by attorneys.[32] Minority defendants receive sentence lengths that are on average .7 months longer, although this effect is not statistically different from 0 at conventional levels. On the other hand, Table 10 reveals that attorneys assign 131 percent larger fines to convicted minority defendants (implied by the coefficient of .927). Thus, the analysis of attorneys' decisions reveals a pattern that is similar to that observed for law students and economics students. Minority defendants are more likely to

[32] Each of the 36 lawyers evaluated all six cases, which yielded 216 decisions of conviction or acquittal (108 cases of white defendants and 108 cases of minority defendants). Fifty-nine of the white defendants were convicted (55 percent), while the conviction rate among minority defendants was 69 percent (74 minority defendants). This implies the existence of 15 excess convicted minority defendants. Ranking the 74 convicted minorities by sentence length and fine shows that 12 defendants received no effective prison term and no effective fine, despite their conviction. These are the marginal convicted defendants, who are arguably not guilty despite being convicted. Of the remaining convicted defendants with effective prison terms of 0 months, we dropped three people who were sentenced to 0 months of effective prison term and the lowest initial fines assigned as punishment (two people with fines of €100 and one person with a fine of €104).

be convicted for the same crime, on the basis of the same arguments made by the prosecutors and the same defense by their attorneys, and they are more likely to receive stiffer punishments on conviction.

We also recruited judges and prosecutors to evaluate the cases. The small size of the judge and prosecutor sample does not allow us to analyze them separately, but we merged them with the attorney sample to investigate the sensitivity of the results. Adding 12 judges or prosecutors to the sample of 36 attorneys generates results similar to those obtained for the sample of attorneys, which suggests that judges' and prosecutors' decisions in these trials are not substantially different from those of the attorneys.[33]

8. Summary and Conclusions

In this paper we investigate whether the decision regarding the guilt or innocence of an individual who is accused of a crime is race blind. We also ask whether the extent of the punishment depends on race. Do the decisions depend on whether the defendant and the evaluator are of the same race? If so, are these influences stronger for white or minority evaluators? If race effects on these decisions exist, are they impacted by whether the evaluator believes that terrorism is a major problem in the country?

Some of these are old questions, and all of them are important for scientific inquiry and public policy. The questions, however, are difficult to answer because of a number of inherent omitted-variables and endogeneity issues. At the heart of the problem lies the near impossibility of creating a counterfactual scenario in a trial that involves the defense attorney, the prosecutor, and the defendant. We create a design that holds constant everything that takes place during a trial, with the exception of the race of the defendant. The VR technology allowed us to alter only the race of the defendants, holding all activity in the courtroom constant, including every word spoken by the prosecutor and the defense attorney and all the body language in the courtroom. This paper is the first to utilize 3D VR technology that uses people rather than computer-animated scenes.

This design allows us to bypass the identification challenges faced by previous empirical research and to create arguably perfect counterfactuals. In our case, the race of the defendant in each trial is uncorrelated with the characteristics of the prosecutors, with the characteristics of defense attorneys, and with any activity in the courtroom. The defendant's race is also uncorrelated with the evaluators' attributes. The only variation in a given trial is the race of the defendant.

---

[33] There is no statistically significant racial bias in prison sentencing in the attorney sample or the sample with attorneys, prosecutors, and judges. The point estimate of Minority Defendant in the conviction decision regression is .139 in the former sample, while it is .103 in the latter sample, and both are significantly different from 0, although the difference in the magnitudes is not inconsistent with the null effect in a sample with only judges and prosecutors. The point estimate of Minority Defendant in the regression for fines is .927 in the attorney sample and .860 in the sample with attorneys, prosecutors, and judges (both significantly different from 0), and the difference in the magnitudes is inconsistent with the hypothesis that the sample of judges and prosecutors would produce a null effect.

The results provide suggestive evidence for negative in-group bias during the conviction phase. Evaluators were more likely to convict a defendant if he was of the same race as the evaluator, although this effect is not statistically significant. Because an evaluator's race is, by design, uncorrelated with all other variables including the race of the defendant, in-group bias can be decomposed to determine its sources. Our analysis shows that the statistically insignificant in-group bias in convictions emerges because of two opposing forces. White evaluators exhibit positive in-group bias, while minority evaluators exhibit negative in-group bias; that is, both white and minority evaluators treat white defendants favorably relative to minority defendants during the conviction stage. These effects produce the overall racial bias against minorities in conviction decisions: minority defendants are 12 percentage points more likely to be convicted than white defendants even though they are tried for the same case and even though everything during the trial is the same.

In the sentencing stage, we find in-group favoritism. Convicted defendants receive prison sentences that are about 4 months shorter if the evaluator is of the same race. We show that this positive in-group bias in prison sentences is driven equally by white and minority evaluators. Because most evaluators are white, this behavior translates into racial bias in prison sentencing, with minority defendants being sentenced to prison terms 1.3 months longer on average, which corresponds to a 32 percent increase in prison time.

Positive in-group bias also exists in the assignment of fines, but this bias is stronger for white evaluators, who assign fines to white defendants that are 68 percent lower than those assigned to minority defendants. This translates into racial bias in fines: minority defendants receive stiffer fines than their white counterparts.

Using the subjective assessments of individuals who were not part of the experiment, we show that controlling for perceived criminal proclivity or perceived honesty of defendants does not alter the results. When we analyze whether disparate treatment of defendants depends on the extent to which evaluators are concerned about terrorism in the country, we find no discernible behavioral difference between those evaluators who are very concerned about terrorism and those who are less concerned. This may suggest that the source of the racial bias is deeply rooted. Alternatively, it may be that the terrorist attacks in Brussels and Paris, which took place about 2 years earlier, generated increased animosity toward minorities that has not yet dissipated and that affects the decisions about minority defendants even though the alleged offenses in these trials are unrelated to any terrorist act.

We repeated the experiment with a group of 36 practicing attorneys, most of whom specialize in criminal law. Although this sample is smaller, it provides stronger external validity. But because all the attorneys in the sample are white, we could not investigate in-group bias in their decisions. The results are similar to those obtained from law students and economics students, both qualitatively and in magnitude. A small group of judges and prosecutors who were not in-

volved in the experiment also participated as evaluators of the VR trials. Their small sample size does not permit us to analyze their decisions separately, but when we merge the sample of judges and prosecutors with the sample of attorneys, the analyses provide results similar to those for the attorney sample, which suggests that judge's and prosecutors' decisions are not substantially different from those of attorneys.

A large number of additional analyses confirm the robustness of the results. For example, the speed with which the evaluators made their decisions or whether the evaluators altered their original conviction or sentencing decisions has no impact on the results. Similarly, the decisions made during the first half of the experiment (the first three trials) are no different from those made in the second half, and the decisions of men and women are similar.

It is important to emphasize that for minorities the in-group bias in the conviction phase is negative but the in-group bias in the sentencing phase is positive. In other words, minority evaluators are harsher toward in-group defendants when it comes to the decision about guilt or innocence, but they favor same-race defendants during the sentencing phase. This finding is important because it reveals a changing pattern of bias (negative then positive in-group bias) in the same decision sequence by the same group of evaluators, and this is the first paper to identify this pattern.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. Do Judges Vary in Their Treatment of Race? *Journal of Legal Studies* 41:347–83.

Agan, Amanda, Matthew Freedman, and Emily Owens. 2021. Is Your Lawyer a Lemon? Incentives and Selection in the Public Provision of Criminal Defense. *Review of Economics and Statistics* 103:294–309.

Alesina, Alberto, and Eliana La Ferrara. 2014. A Test of Racial Bias in Capital Sentencing. *American Economic Review* 104:3397–3433.

Anderson, James M., and Paul Heaton. 2012. How Much Difference Does the Lawyer Make? The Effect of Defense Counsel on Murder Case Outcomes. *Yale Law Journal* 122:154–217.

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia. *American Economic Review* 96:847–62.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. The Impact of Jury Race in Criminal Trials. *Quarterly Journal of Economics* 127:1017–55.

Argys, Laura M., and H. Naci Mocan. 2004. Who Shall Live and Who Shall Die? An Analysis of Prisoners on Death Row in the United States. *Journal of Legal Studies* 33:255–82.

Arnold, David, Will Dobbie, and Peter Hull. 2020. Measuring Racial Discrimination in Bail Decisions. Working Paper No. 26999. National Bureau of Economic Research, Cambridge, MA.

Ayres, Ian, and Peter Siegelman. 1995. Race and Gender Discrimination in Bargaining for a New Car. *American Economic Review* 85:304–21.

Bertrand, Marianne, and Esther Duflo. 2017. Field Experiments on Discrimination. Pp.

309–93 in *Handbook of Field Experiments,* edited by Abhijit Vinayak Banerjee and Esther Duflo. Amsterdam: North-Holland.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *American Economic Review* 94:991–1013.

Chen, Yan, and Sherry Xin Li. 2009. Group Identity and Social Preferences. *American Economic Review* 99:431–57.

Depew, Briggs, Ozkan Eren, and Naci Mocan. 2017. Judges, Juveniles, and In-Group Bias. *Journal of Law and Economics* 60:209–39.

Eren, Ozkan, and Naci Mocan. 2018. Emotional Judges and Unlucky Juveniles. *American Economic Journal: Applied Economics* 10:171–205.

Gazal-Ayal, Oren, and Raanan Sulitzeanu-Kenan. 2010. Let My People Go: Ethnic In-Group Bias in Judicial Decisions—Evidence from a Randomized Natural Experiment. *Journal of Empirical Legal Studies* 7:403–28.

Glaeser, Edward L., and Bruce Sacerdote. 2003. Sentencing in Homicide Cases and the Role of Vengeance. *Journal of Legal Studies* 32:363–82.

Goette, Lorenz, David Huffman, and Stephan Meier. 2006. The Impact of Group Membership on Cooperation and Norm Enforcement: Evidence Using Random Assignment to Real Social Groups. *American Economic Review* 96:212–16.

Grossman, Guy, Oren Gazal-Ayal, Samuel D. Pimentel, and Jeremy M. Weinstein. 2016. Descriptive Representation and Judicial Outcomes in Multiethnic Societies. *American Journal of Political Science* 60:44–69.

Heckman, James J. 1998. Detecting Discrimination. *Journal of Economic Perspectives* 12(2): 101–16.

Heckman, James, and Peter Siegelman. 1993. The Urban Institute Audit Studies: Their Methods and Findings. Pp. 187–258 in *Clear and Convincing Evidence: Measurement of Discrimination in America,* edited by Michael E. Fix and Raymond J. Struyk. Washington, DC: Urban Institute Press.

Hoffman, Florian, and Philip Oreopoulos. 2009. A Professor like Me: The Influence of Instructor Gender on College Achievement. *Journal of Human Resources* 44:479–94.

Jones, Angela M., Amanda N. Bergold, Marlee Kind Dillon, and Steven D. Penrod. 2017. Comparing the Effectiveness of *Henderson* Instructions and Expert Testimony: Which Safeguard Improves Jurors' Evaluations of Eyewitness Evidence? *Journal of Experimental Criminology* 13:29–52.

Kennedy, Peter E. 1981. Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations. *American Economic Review* 71:801.

Levitt, Steve D. 2004. Testing Theories of Discrimination: Evidence from *Weakest Link*. *Journal of Law and Economics* 47:431–52.

List, John A. 2004. The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field. *Quarterly Journal of Economics* 119:49–89.

List, John A., and Michael K. Price. 2009. The Role of Social Connections in Charitable Fundraising: Evidence from a Natural Field Experiment. *Journal of Economic Behavior and Organization* 69:160–69.

Maeder, Evelyn M., Annik Mossière, and Liann Cheung. 2012. Candian Mock Juror Attitudes and Decisions in Domestic Violence Cases Involving Asian and White Interracial and Intraracial Couples. *Journal of Interpersonal Violence* 28:667–84.

Mendoza, Saaid A., Sean P. Lane, and David M. Amodio. 2014. For Members Only: In-group Punishment of Fairness Norm Violations in the Ultimatum Game. *Social Psycho-*

logical and Personality Science 5:662–70.

Mercado, Cynthia Calkins, Brian H. Bornstein, and Robert F. Schopp. 2006. Decision-making about Volitional Impairment in Sexually Violent Predators. *Law and Human Behavior* 30:587–602.

Mossière, Annik, Evelyn M. Maeder, and Emily Pica. 2018. Racial Composition of Couples in Battered Spouse Syndrome Cases: A Look at Juror Perceptions and Decisions. *Journal of Interpersonal Violence* 33:2867–90.

Neumark, David, Roy J. Bank, and Kyle D. Van Nort. 1996. Sex Discrimination in Restaurant Hiring: An Audit Study. *Quarterly Journal of Economics* 111:915–41.

Ondrich, Jan, Stephen L. Ross, and John Yinger. 2000. How Common Is Housing Discrimination? Improving on Traditional Measures. *Journal of Urban Economics* 47:470–500.

Pager, Devah. 2003. The Mark of a Criminal Record. *American Journal of Sociology* 108:937–75.

Pager, Devah, Bruce Western, and Naomi Sugie. 2009. Sequencing Disadvantage: Barriers to Employment Facing Young Black and White Men with Criminal Records. *Annals of the American Academy of Political and Social Science* 623:195–213.

Rehavi, M. Marit, and Sonja B. Starr. 2014. Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy* 122:1320–54.

Riach, P. A., and J. Rich. 2002. Field Experiments of Discrimination in the Market Place. *Economic Journal* 112:F480–F518.

Shayo, Moses, and Asaf Zussman. 2011. Judicial Ingroup Bias in the Shadow of Terrorism. *Quarterly Journal of Economics* 126:1447–84.

Turner, Margery Austin, Michael Fix, and Raymond J. Struyk. 1991. *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring.* Washington, DC: Urban Institute Press.