

Racial Bias in the Criminal Justice System

Jennifer L. Doleac
Texas A&M University

June 2021

Accepted for publication in: "A Modern Guide to Economics of Crime," edited by Paolo Buonanno, Juan Vargas, and Paolo Vanin. Edward Elgar Publishing.

Introduction

Racial disparities have been a prominent feature of the U.S. criminal justice system for many decades, and these persistent disparities are of concern to policymakers, advocates, and citizens. In 2019, 26.6% of individuals arrested were Black (FBI UCR, 2019), and 32.8% of sentenced inmates were Black (Carson, 2020). However, only 13.4% of U.S. residents are Black (Census, 2019).

To design effective policies to reduce racial disparities in criminal justice outcomes, it's important to understand what is driving them. It's possible that a large share of existing disparities in criminal justice involvement is caused by differences in actual criminal behavior across groups — perhaps driven by differences in poverty or access to education. In these cases, addressing those root causes of differences in behavior will be required to reduce racial disparities in criminal justice outcomes.

However, it is also possible that disparate treatment by the criminal justice system — that is, racial bias from police, prosecutors, and judges — contributes to existing disparities in outcomes and should be addressed directly.¹ And when evidence of racial bias exists, understanding whether that bias represents animus (bias based on race itself), statistical discrimination (using race as a proxy for some other, unobservable, factor), or stereotyping (statistical discrimination based on inaccurate information) is similarly important for designing policies that can reduce that bias (see Doleac, 2020, for a full discussion and examples).

In this chapter I discuss the current economics literature on racial bias at various stages of the criminal justice system (policing, prosecution, trial/sentencing, reentry). I also discuss what we currently know about what types of interventions work to reduce this bias. Throughout the chapter, I highlight the research frontier -- key questions of interest for those concerned about racial bias in the criminal justice system and how to reduce it.

Evidence on the existence of racial bias

Policing

Police interactions with civilians come about in two ways: When on patrol, police might choose to stop a particular person (pedestrian or driver). Alternatively, they might be sent to a

¹ It's also possible that Black and white Americans are subject to different criminal justice policies because they live in different places. There is evidence that racial bias drives electoral preferences about the punitiveness of local criminal justice policies (Feigenberg and Miller, 2021). In this case, Black and white residents of a particular place are not necessarily treated differently, but differences in policies across place contribute to racial disparities in outcomes at the national level. See Albright (2019) for evidence of this in the context of pretrial detention.

particular place in response to a 911 call for service. In each case, the police officer has a lot of discretion over how to handle the incident after the initial interaction occurs; the end result might be a warning or citation, an arrest, or even violence.

An important question is whether racial disparities in these outcomes are due to racial bias by officers, rather than differences in behavior by civilians. This question is difficult to answer because the researcher typically does not see everything that the officer saw in the moment — in particular, the civilian's behavior. Even obtaining detailed police reports does not solve this problem. Information recorded in police reports might — intentionally or not — be slanted to make the officer's response appear more justified than it was (e.g. claiming "furtive movements" that seemed suspicious or, in extreme cases, lying about evidence that was planted by the officer). This would bias any analyses that condition on this information toward finding no racial bias.

A large literature attempts to address the question of racial bias in choosing whom to stop, particularly in traffic stops. The so-called "veil of darkness" test developed by Grogger and Ridgeway (2006) compares individuals stopped during daylight and at night, with the intuition that it is more difficult to see drivers' race at night when it is dark. If there is a difference in the racial composition of stops just before and after the sun sets, but all other observable characteristics are the same, this is evidence that officers considered race when deciding whom to stop during the day. This study uses Daylight Saving Time as a natural experiment to compare the same hour of the day before and after dark; this holds other factors such as commuting patterns constant. Using data from Oakland, CA, they do not find evidence of racial bias in traffic stops.

Horrace and Rohlin (2016) note that many urban areas are well-lit even at night, and so the veil-of-darkness test will be more useful in areas without street lighting. They use data from Syracuse, NY, that included streetlight locations, and find that Black drivers are disproportionately more likely to be stopped during the day than at night when it is dark. This implies racial bias against Black drivers in the decision of whom to stop.

Another influential strain of this literature uses "outcome tests" (Becker, 1957; Knowles, Persico, and Todd, 2001) to test for racial bias in the choice of whom to stop (both pedestrians and drivers). The intuition for these tests is that officers should be stopping people based on their probability of some misconduct — for instance, carrying contraband. They should therefore choose to stop people from various groups in a way that maximizes the likelihood of achieving some outcome (finding contraband, or making an arrest). This implies that that marginal white person who is stopped and the marginal Black person who is stopped should have equal probabilities of carrying contraband. If the probability of finding contraband on the marginal white person is higher, this implies that officers are acting inefficiently and they should stop more white people and fewer Black people. If they don't adjust in this way, this implies racial bias against the group with the lower "success" rate.

Antonovics and Knight (2009) refine this test by using the interaction of officer and driver race to determine whether racial bias is due to animus or statistical discrimination. The intuition behind this approach is that if statistical discrimination is driving behavior toward Black drivers, then it should be the same for white and Black officers alike. (This is because all officers are incentivized to use information that helps them infer criminality; if a driver's race is valuable in this way, then white and Black officers will both use it.) However, animus against Black drivers is less likely from Black officers, because individuals are unlikely to exhibit animus toward their own race.

A challenge in this literature is identifying the marginal person stopped — the so-called "inframarginality problem." The original theory papers that devised this test essentially assumed

representative agents where the marginal person was the same as the average person.² They could thus compare averages across groups to test for racial bias. But of course in the real world people are not homogenous, and the marginal person will typically not be the same as the average.³ So, while outcome tests are appealing intuitively, and have been influential in policy settings, they are problematic in practice.

There is an active literature working to refine these tests and clarify their implications (see for example: Canay, Mogstad, and Mountjoy, 2020; Gelbach, 2021; and Hull, 2021). These studies highlight (at least) two additional issues with outcome tests: (1) Outcome tests may fail — thus implying racial bias — due to “omitted payoff bias.” That is, the outcome that is the focus of the test is not the (only) outcome that officers care about. For instance, officers may stop drivers in the hope of finding contraband *or* arresting people with outstanding warrants. An outcome test based only on the former might fail, but might pass if both outcomes were considered. In practice, it is difficult for researchers to know officers’ complete objective functions. (2) Statistical discrimination may be based on inaccurate information (“stereotypes” in the terminology of Bordalo, et al., 2016).

Both of these issues further complicate the value of outcome tests. Hull (2021) shows that omitted payoff bias is empirically indistinguishable from the use of stereotypes, though of course the two possibilities have very different implications (the first may not be concerning, but the second would be unconstitutional).

Another problem with outcome tests in practice is that incentivizing officers to equalize “success” rates across groups could push racially-biased officers to lie about the outcome of the search. They might plant contraband on Black drivers to make those stops look successful based on the metrics of interest, or misreport other details of the incident. Indeed, Luh, 2020, shows that officers in Texas systematically misreported Hispanic drivers as white when searches failed, in order to make themselves look less biased. Thus, while outcome tests might be useful in a vacuum, they can make problems worse if used as key performance metrics in the real world.

More recent work tests for racial bias in police behavior, separate from the initial decision to stop someone. Jeremy West (2018) uses car crashes in a large unnamed state as a natural experiment. In this context, the closest officer is dispatched to the scene to provide aid and address any wrong-doing. The officer has discretion over how to handle the incident, but does not have discretion over which drivers to engage with. West shows that, after controlling for the local geographic area, the race of the driver(s) involved in the crash and the race of the officer dispatched to the scene appear random. This allows him to test for racial bias by the officers based on the race of the drivers. He found substantial evidence of in-group bias (e.g., white officers favored white drivers), particularly for low-level decisions over which officers have the most discretion — whether to cite someone for an expired registration for instance. West does not find evidence of bias in more serious outcomes — whether to charge someone with a felony offense — presumably because there is more oversight there. Since the low-level offenses typically involve clear evidence of wrong-doing (the driver’s registration is either expired or it’s

² An exception is the rank-order test developed by Anwar and Fang (2006). This test considers whether the ranking of driver search rates (by race group) varies with officer race. If it does, this implies racial animus (versus statistical discrimination or no discrimination) on the part of officers.

³ This said, Feigenberg and Miller (2020) suggests that at least in one context the average and marginal hit rates appear to be similar. More work on this, using changes in the number of people stopped to identify those at the margin, would be helpful.

not, there's no need to guess), West argues that the bias he measures is likely racial animus rather than statistical discrimination.

Emily Weisburst (2017) uses a similar scenario to test for racial bias in the outcomes of calls for service in Dallas. She argues that unlike voluntary interactions on patrol, officers have no discretion over which calls for service to respond to. When a 911 call comes in, officers are dispatched to the scene based on who is the nearest available officer. As in West (2018), once she controls for the geographic area the race of the officer and the race of the civilian at the scene are essentially random. She tests for racial bias in the likelihood that officers make arrests when dispatched to a call involving an other-race civilian. While there is a great deal of variation across officers in the likelihood that they make arrests (highlighting the amount of discretion they have), Weisburst does not find evidence of racial bias in this setting. Hoekstra and Sloan (2020) use the same strategy in two unnamed cities, and find evidence of racial bias in officers' use of force against civilians. The different result in that paper might be due to the different outcome measure (force versus arrests), or the different context — there is surely tremendous variation in officers' behavior across cities.

Another study of note in this area is Goncalves and Mello (2020). Using data from Florida, they note a discontinuity in punishment for speeding, at 10 miles per hour (mph) over the speed limit. Those recorded as driving 9 mph over the limit receive a warning, while those recorded as driving 10mph over the limit receive a monetary fine. However, officers have discretion in what speed they say the driver was driving. Goncalves and Mello test for discontinuities in the likelihood of being recorded as driving 9 versus 10 mph over the limit, and find that white drivers are much more likely to be recorded as driving 9mph over the limit and let off with a warning. This suggests racial animus on the part of officers against Black drivers.

All of these papers use natural experiments in clever ways to test for differences in outcomes based on race. One shortcoming in each case, however, is that the researchers are still unable to observe the drivers'/civilians' behavior. While the race-match of the officer and civilian is random, and it is plausible that the pre-interaction characteristics of the white and Black drivers are the same (given the natural experiment that produced the interaction), once the officer arrives at the scene the behavior of both the officer and the civilian, and how they respond to one another, produce the outcome (a citation, arrest, or violence). The papers described above implicitly assume that differences in the officer's behavior is what drove any observed differences; this is a plausible assumption given the power that officers have in these contexts. But it is possible that civilians' behavior is also different, perhaps in response to the race of the officer. Ultimately what these papers are measuring is the net result of the officer-civilian pair.

Pretrial Detention

After someone is arrested, they might be held in jail while awaiting trial. In many cases, cash bail may be set, and at least a fraction of that amount would need to be paid to secure the defendant's release. In other cases, bail judges or magistrates may simply decide to hold a defendant in pretrial detention because they are a flight risk or threat to public safety. A number of studies have found that, for those on the margin of pretrial detention, being held in jail while awaiting trial has substantial detrimental effects: it increases the likelihood of conviction in the current case, as well as subsequent criminal justice involvement. It also increases court debt and reduces future employment. (See Heaton et al., 2017; Leslie and Pope, 2017; Dobbie et al., 2018; and Stevenson, 2018.) So, racial bias in the pretrial detention decision could exacerbate racial disparities in criminal justice outcomes as well as employment.

Arnold, Dobbie, and Yang (2018) use the quasi-random assignment of defendants to bail judges, along with the incidence of pre-trial misconduct among those who are released, to test for racial bias in pre-trial detention in Miami and Philadelphia. They find that black defendants are more

likely to be detained pre-trial than similar white defendants are. The authors argue that this is evidence of racial bias against black defendants based on inaccurate stereotypes of their criminality.

Note that focusing on the incidence of pre-trial misconduct as an indicator of whether someone should have been released is an outcome test, with all the problems and caveats discussed above. In this context, because pretrial detention is a function of the judge's decision to set cash bail and the defendants' ability to pay that bail, it is possible that the higher rate of pre-trial detention for Black defendants is due to judges' not adjusting the bail amount (enough) based on those defendants' ability to pay. That is, if they set equal bail for similar Black and white defendants, but Black defendants are poorer and less able to pay the required amount, pretrial detention rates will be higher for Black defendants. This makes it more difficult than in other contexts to be sure that racial bias is driving the observed effect. (This possibility is also raised in Gelbach, 2021, as one reason that the data from this study appear inconsistent with basic implications of the Becker model of discrimination.)⁴

Prosecution

After the police issue a summons or arrest a suspect, a prosecutor decides whether to move forward with the case. There has been a great deal of recent attention to the discretion that prosecutors have in determining the outcomes of individual cases (see for example, Pfaff, 2017, and Agan, Doleac, and Harvey, 2021); this discretion makes room for racial bias to affect decisions.

A seminal paper in this space is Rehavi and Starr (2014). Using data on federal cases from arrest through sentencing, they control for a rich set of observable case characteristics and find that the race of the defendant explains substantial remaining variation in case outcomes. They show that most of the racial differences in outcomes are driven by racial differences in prosecutors' initial charging decisions; Black defendants are substantially more likely to face a charge that carries a mandatory minimum sentence than similar white defendants are.

Of course, there may be unobservable differences across cases with Black and white defendants that the researchers were unable to control for and that may be driving the case outcomes. Sloan (2020) uses the randomization of cases across prosecutors in New York County to test for racial bias in case outcomes. She finds evidence of in-group bias in property offense cases (e.g., white prosecutors are more lenient toward white defendants), driven by prosecutors' decisions to dismiss charges in some cases but not others.

Yang (2015) considers the effects of *United States v. Booker*, which struck down federal sentencing guidelines. Overall she found that this change led to an increase in racial disparities in sentences (more on this below), but part of this effect was driven by prosecutors' charging decisions: once sentencing guidelines were removed, prosecutors became more likely to charge Black defendants with offenses that carried mandatory minimum sentences. This increased the gap between sentences received by white and Black defendants.

Along similar lines, Tuttle (2021) uses a federal policy change related to drug offenses as a natural experiment. He shows that when the mandatory minimum threshold for crack-cocaine increased from 50g to 280g, prosecutors responded by increasing the fraction of cases sentenced at 280g, presumably to qualify for the mandatory minimum sentence. This increase was more likely for Black and Hispanic defendants. Tuttle argues that this reflects prosecutorial discretion

⁴ Also see Arnold et al. (2020) for an extension of this method in the pre-trial context.

in the charging decision, and also shows that differences in prosecutors' responses are correlated with state-level measures of racial animus.

Conviction and Sentencing

There is a large literature on the effect of individual judges on sentencing decisions, exploiting quasi-random variation in event timing or assignment of cases to courtrooms. Previous work has demonstrated that sentencing decisions — by experienced judges — are routinely affected by irrelevant information such as temperature (Heyes and Saberian, 2019), media coverage of crime (Philippe and Ouss, 2018), and the order in which cases are heard (Chen, Moskowitz, and Shue, 2016).

It seems likely that this variation will disproportionately harm disadvantaged defendants, including Black defendants. For instance, Eren and Mocan (2018) find that when a local football team unexpectedly loses a game (what the authors describe as an emotional shock), judges issue harsher sentences during the following week. This effect is largest for Black defendants.

There is other evidence that race itself affects judicial decisions. Abrams, Bertrand, and Mullainathan (2012) use random variation of felony cases across judges in Cook County, IL, to consider racial gaps in sentence outcomes for similar Black and white defendants. They find significant differences across judges in these racial gaps when looking at incarceration rates. This implies that some judges are paying attention to race when deciding whom to send to prison.⁵

Alesina and La Ferrara (2014) use rates of sentence reversal in cases where defendants were initially sentenced to capital punishment, to test for racial bias in those initial sentences. They find that the sentences of minority defendants accused of killing white people are much more likely to be overturned upon appeal. This effect is confined to Southern states. The authors interpret this result as implying racial bias against minority defendants in those states.

There is surely variation across contexts in the presence and extent of such bias. For instance, Depew, Eren, and Mocan (2017) find evidence of negative in-group bias toward juvenile defendants by judges in an unnamed US state. Juvenile defendants who are the *same* race as the judge (that is, white defendants in white judges' courtrooms) are more likely to be incarcerated than similar defendants of another race.

Park (2017b) uses administrative data on sentencing decisions in Kansas to test the hypothesis that judges become tougher on crime (sentence more harshly) during the weeks and months leading up to an election. He finds strong, robust evidence that sentencing increases during this period, but only for black defendants, and only in places with partisan judicial elections. Park argues that this change in judicial sentencing behavior is due to electoral pressures on judges themselves. These effects are larger for Democratic judges (who might be more sensitive to “soft on crime” attacks) and in places where (based on a couple of different measures) voters are more racially biased.

Though relatively few cases in the United States go to trial, juries still play a role in case outcomes. And race matters there, too. Anwar, Bayer, and Hjalmarsson (2012) use data from Florida and random assignment of local residents to jury pools to show that juries formed from

⁵ But also see Park (2017a), which applies Anwar and Fang's (2006) rank-order test to judicial sentencing to determine whether racial animus is at work. Using data from Kansas, he cannot reject the null hypothesis of no racial animus in judicial decision-making.

all-white jury pools are substantially more likely to convict black defendants than are juries formed from jury pools with at least one black member.

Reentry

After an individual leaves prison, they must reenter civilian society. A key step is finding non-criminal employment. There is extensive evidence that employers are reluctant to hire people with criminal records; there is also extensive evidence that employees discriminate against Black applicants. A more open question is how race and criminal records interact in the labor market.

Pager (2003) conducted an audit study where Black and white individuals applied for jobs in person, randomizing whether they said they had a criminal record or not (but keeping all other qualifications the same). She found that not only did employers discriminate against people with criminal records, but that discrimination was more pronounced for Black applicants. In a striking result, Pager showed that, at least in that context, white applicants with a criminal record received more callbacks from employers than did Black applicants without a criminal record.

A more recent audit study by Agan and Starr (2018) also found that employers discriminate against people with criminal records, but did not find that this effect varied by race. In fact, callback rates were nearly identical for white and Black applicants with the same criminal record status. This different result might be due to the different time period, or the different place (New Jersey and New York).

Most people who are convicted of a crime will spend some time on community supervision, either on probation or parole. During this period, breaking the rules of that supervision — “technical violations” of probation/parole, such as being out past curfew, missing a meeting, or failing a drug test — can result in incarceration. Probation and parole officers have substantial discretion in when to write someone up for a technical violation. Rose (2020) considers who is incarcerated for technical violations of supervision, in the context of probation in North Carolina. He finds that technical rules disproportionately affect low-risk Black probationers. Ostensibly race-neutral rules related to the payment of fees and fines drive much of this racial disparity in who is incarcerated due to technical violations of probation.

What works — and what doesn't — to reduce bias

There are many possible ways to reduce bias in decision-making, but we have very little evidence on what works. Below I discuss a few possible approaches, and the current evidence (if any) on their efficacy.

Changing the composition of decision-makers

It is possible that the current set of decision-makers in the criminal justice system is more biased than others might be in the same role. If so, changing who holds these positions could reduce bias in the decisions made.

Studies such as Hoekstra and Sloan (2021) and Ba, et al. (2021) provide useful descriptive evidence that Black and white officers respond differently to crime incidents involving Black residents. But such descriptive evidence does not tell us whether hiring more Black officers would be helpful — the marginal Black (white) officer hired might be different from the average Black (white) officer currently on the force. To know whether changing the composition of police forces would have beneficial effects, we need an intervention that changes the composition of the force.

There is some evidence that increasing the number of Black officers on police forces reduces arrests of Black residents (McCrary 2007), with no detrimental effects on crime rates; similarly, increasing the number of women officers reduces violence against women (Miller and Segal, 2019). These studies use court-ordered affirmative action as natural experiments. These court orders increased the share of Black and women officers in police departments, allowing researchers to measure the effects of such a change on other outcomes.

Harvey and Mattia (2021) show that litigation against police departments in response to employment discrimination appears to have reduced discriminatory behavior throughout the institution, including among rank-and-file officers. They find that after such litigation is filed (even before the composition of the police force changed), Black-white disparities in crime victimization fell, and Black residents became more likely to report crimes to police (suggesting increased trust in the police to help them).

Of course, simply saying we want a more diverse police force (or other criminal justice agency) does not make it so. Current recruitment strategies may not be effective at encouraging new and different people to join the police. Linos (2018) describes an experiment that tested different recruitment messages in Chattanooga, TN. She found that the standard messaging — about public service — had no effect on applications. However, messages emphasizing that policing is a challenging job and a long-term career increased applications, particularly from underrepresented groups. The next step in this research space will be to see if the different people hired as a result of these different messages behave in a less-biased manner on the job.

Existing rules about who is eligible to become a police officer may unnecessarily reduce diversity. Many departments require that applicants have no criminal record, have no tattoos, have a college degree, and/or not use marijuana (even if it is legal in the state). It would be useful to measure the effects of such rules on hiring as well as on policing outcomes.

There is much less work on how the composition of other criminal justice agencies affects racial disparities in outcomes. Cohen and Yang (2019) show that federal judges appointed by Republicans issue relatively harsh sentences to Black (vs. white) and male (vs. female) defendants, when compared with judges appointed by Democrats. However, as discussed above, Park (2017b) shows that judges appear to respond to the demands of their constituency — that is, they are harsher toward Black defendants leading up to an election in places where the local electorate is more racially biased. This highlights the role that the local electorate plays in determining the behavior of those who have decision-making authority.

Training

It is possible that training decision-makers to recognize their biases and change their behavior might reduce racial discrimination. Implicit bias training is often used to try to achieve this goal, but there is currently no rigorous evidence that it is effective at changing behavior (Green and Hagiwara, 2020).

Procedural justice training could reduce disparities in outcomes if they improve trust and communication between police and civilians. Owens et al. (2018) find that such training has beneficial effects, in a relatively small experiment in Seattle. Wood et al. (2020) and Roth & Sant'Anna (2021) present results from a large-scale experiment evaluating a one-day procedural justice training for police officers in Chicago. In general, this one day training does not appear to have had effects on citizen complaints or police use of force.

It is possible that other types of training could be effective, either for new hires or as an intervention for those who have exhibited bias on the job, but we do not yet have any evidence

that this approach works. The staggered rollout model from Wood et al. (2020) and Roth and Sant'Anna (2021) will surely be useful for testing other types of trainings in the field.

Change the amount of information provided

One common approach to reduce discrimination based on particular information (such as a criminal record) is to remove that information. But this can increase statistical discrimination against a broader group that is more likely to have that characteristic. See Doleac, 2021, for a full discussion.

One relevant example is Ban the Box policies, which prevent employers from asking about a job applicant's criminal record until late in the hiring process. The goal of these policies is to help people with records get their foot in the door and increase the likelihood that they'll be hired for the job. But if employers don't want to hire people with records, and this policy prevents them from asking who has a record, then they now have an incentive to guess. This could lead them to discriminate against young, Black men without college degrees — the group that is most likely to have a recent conviction that might worry an employer. Indeed, Agan and Starr (2018) and Doleac and Hansen (2020) show that Ban the Box policies do increase statistical discrimination against young Black men, which reduces employment for those without records and leads to a net decline in employment for this group. Sherrard (2020) shows that this leads to an increase in recidivism for Black people with a previous conviction. Thus, Ban the Box policies appear to increase, rather than reduce, racial disparities in employment and criminal justice outcomes.

Expunging or clearing records could have similar unintended consequences, if it removes information that employers care about, thus leaving them to guess who has a criminal history that they now can't see. However, if employers are most concerned about the legal liability associated with hiring someone with a criminal record, then clearing records could be helpful — since such records aren't visible to employers, those employers can't be accused of negligent hiring if the person later commits a crime on the job. Since the effects of these policies will hinge on why exactly employers discriminate against people with criminal records — which we don't yet fully understand — it will be important to evaluate their real-world effects. Doleac and Lageson, 2020, discuss the various reasons that popular record-clearing policies might fail, and call for more research on this topic.

Since removing information often has important unintended consequences, it may be more productive to *add* information. This will be a useful strategy if the type of discrimination at work is statistical discrimination — where decision-makers use race (or something correlated with race, such as a criminal record) to infer other information they can't see, such as likelihood of reoffending, or productivity on the job. If we can provide better information about the characteristics that decision-makers are trying to guess, they'll have less need to rely on race.

In the employment context, court-issued rehabilitation certificates seem promising for this reason (Leasure and Stevens Andersen, 2016; Leasure and Martin, 2017). Completing a rigorous job-training or rehabilitation program could also send useful signals to employers about an applicant's reliability and work-readiness; in line with this idea, there is anecdotal evidence of such programs becoming feeders for local employers, who trust that program graduates will make good employees (Piehl, 2009).

Providing more information is also a motivation for risk assessment scores. Such scores are now used throughout the criminal justice system to inform decisions. They use relevant information to predict someone's risk of misconduct or reoffense, in a standardized way across all individuals. Thus, white and Black defendants accused of the same crime and with the same criminal history, age, gender, and so on would have the same risk score. This might be helpful if judges are inclined to view the Black defendant as higher risk due to racial bias. Seeing the

identical risk scores across similar white and black defendants could push them to treat the two defendants equally, if they believe that the risk scores provide more accurate measures of true risk than they (the judges) could infer on their own.

Researchers have been particularly interested in whether providing these scores to judges who make decisions about pre-trial detention and sentencing might improve those decisions (targeting incarceration more efficiently) and reduce racial disparities in outcomes. Policy simulations suggest that replacing judges' decisions with the outcome of the risk assessment would have beneficial effects (Kleinberg et al., 2018), but in practice this policy does not seem effective. Judges simply don't use the additional information, or use it in a racially-biased manner — for instance, treating white and Black defendants with the same score differently (Stevenson 2019; Stevenson and Doleac, 2021). The result is no net reduction in racial disparities within a place.⁶

There is also evidence that risk assessment is adopted and used in disparate ways across places, with the end result being that this policy can actually increase, rather than reduce, racial disparities in outcomes. (This is in line with Feigenberg and Miller, 2021, which shows that racial disparities can be driven by harsher policies being implemented in more racially diverse places). For instance, in Kentucky, it appears that risk assessment changed judges' behavior most in counties with more white residents. White residents of the state thus benefited disproportionately from the lower pretrial detention rates encouraged by the risk scores, widening racial disparities in pretrial detention at the state level (Albright 2019).

One reason that risk assessment tools may not be having the big benefits advocates hoped for is that they do not provide additional information that judges find useful. Judges might think they are just as good at determining risk as the risk score is, but risk is not the only factor they care about. In particular, judges may be releasing people who appear high-risk because of mitigating circumstances such as age (Stevenson and Doleac, 2021). Youth is a highly-predictive risk factor for misconduct and criminal activity, but is widely considered a reason for leniency by criminal justice decision-makers.

That said, we know that human decision-makers are easily distracted by irrelevant information, including race. Pushing them to recognize these errors and biases — with standardized risk scores or something else — still holds promise. But existing studies on the real-world effects of these tools show that we have a long way to go to figure out how to implement such policies in a way that achieves our goals.

Limit discretion in decision-making

Human decision-makers are biased, and when they have more discretion there is more opportunity for those biases to affect outcomes. One way to reduce racial disparities in outcomes, then, is to limit discretion in decision-making.

⁶ The good news is that the use of risk scores also don't seem to increase racial disparities. Racial bias is "baked in" to the data used by risk assessment algorithms — for instance, if Black men are arrested and convicted at higher rates than white men who committed the same offense, they will appear higher risk to the algorithm. Many worry that the resulting risk scores will thus necessarily push judges to treat Black defendants more harshly. But the appropriate counterfactual isn't someone's true risk level, it's the judge's perception of their risk. If judges themselves are using the same data, the risk score might not be worse than the judge's perception. And if the judge is biased in their interpretation of that data, then the risk score could be *less* biased than the judge. The policy-relevant question is whether the use of risk scores leads to larger or smaller racial disparities in outcomes in practice.

We see this in sentencing guidelines for those convicted of crimes: a range of acceptable sentences given the defendant's current offense and criminal history. As mentioned above, Yang (2015) measures the effects of *United States v. Booker*, which struck down federal sentencing guidelines. She found that this change — which gave judges and prosecutors more discretion in determining the appropriate sentence — resulted in an increase in racial disparities in sentences.

In theory, risk assessment scores could work the same way, by pushing judges to make the same decision for defendants with the same risk score. But Stevenson and Doleac (2021) find that this does not happen in practice. This is likely because there is no penalty to ignoring the risk scores — if judges were required to treat defendants with the same risk score the same way, we might see a reduction in racial disparities.

West (2018) found that, in the policing context, racial disparities were largest on outcomes where there was limited oversight (e.g. whether to write someone up for an expired registration). That is, racial disparities were wider when police officers had more discretion. Finding ways to reduce police discretion is difficult but would likely reduce racial disparities in outcomes. For instance, it is possible that requiring police officers to wear body cameras and record all interactions with civilians might push them to treat people with the same observed behavior (e.g. an expired registration, shown on camera) the same way. However, we don't yet have any evidence on this.

Prosecutors also have a great deal of discretion about whether to pursue particular charges against a defendant. In addition to sentencing guidelines, which limit the range of penalties they can seek for a particular charge, restricting their ability to pursue particular charges in the first place could be helpful. Particularly for lower-level offenses like nonviolent misdemeanors, where there is lots of variation in whether prosecutors dismiss the charges or pursue a conviction (Agan, Doleac, and Harvey, 2021), instituting clear office policies about how to handle such cases would likely reduce racial disparities. At the legislative level, decriminalizing particular offenses (such as minor drug possession) achieves this goal. Within prosecutors' offices, a policy that all prosecutors will decline to prosecute particular charges would reduce (or even eliminate) racial disparities in outcomes for those charges. Of course, such policies might have other effects (reducing deterrence and increasing the incidence of those offenses), so equity concerns should be weighed against the possibility of an increase in that (perhaps undesirable) behavior. Since decriminalization and decline-to-prosecute policies are increasingly popular in many U.S. cities and counties, more research on these tradeoffs would be helpful.

Increase oversight/accountability

Limiting discretion will often require greater oversight and accountability, perhaps from outside the agency of interest. In the case of policing, civilian oversight boards could serve this role, though it is unclear how effective they are in practice. Ba (2020) shows that making it easier for civilians to file complaints about police misconduct has beneficial effects.

In general it seems likely that increasing the availability of data on police conduct and outcomes would help reduce racial disparities by making those disparities salient to the local community. Luh (2020) highlights that ensuring the accuracy of those data is also important. As discussed above, she found that when police were asked to record their best guess as to a civilian's race, they systematically misreported Hispanic drivers as white when a search did not turn up any contraband. This made the officers look less racially biased than they were. A policy change that required officers to ask drivers their race appears to have corrected this problem. In addition, officers that the new data revealed to be racially-biased faced professional consequences: they were less likely to be promoted or receive raises. Thus, this simple change in how race data were

collected and reported provided a meaningful incentive to officers to behave in a less-biased manner.

As discussed above, external pressure from the judiciary can also be important. Harvey and Mattia (2021) show that court litigation addressing discrimination in police employment appears to have reduced discriminatory behavior throughout the department, thus reducing racial disparities in crime victimization. Similarly, Rivera and Ba (2019) showed that court decisions that increased the likely consequences for police misconduct — decisions that were made salient by memos from union leadership — had big benefits in terms of reducing citizen complaints against police (implying less police misconduct) without increasing crime rates.

To the extent that racial bias by police results in lawsuits about police misconduct, changing who pays for legal settlements or requiring that individual officers carry misconduct insurance (similar to malpractice insurance for doctors) could align incentives to make sure that police unions, leadership, and officers change their behavior. Currently, these individuals and entities are typically protected from any financial consequences of such misconduct, which means they have no incentive to reduce racial bias (Schwartz, 2020).

Discussion

At this point there is a great deal of evidence that racial bias is a problem in essentially all areas where human decision-making is at work; it is not surprising, then, that it is also a problem in the criminal justice system. As discussed above, there is evidence of racial bias at every stage of the criminal justice process, though there is variation in the presence, extent, and type of bias (animus versus statistical discrimination) in different contexts.

The primary research frontier at this point is not identifying the presence of racial bias, but figuring out how to reduce the bias that exists. That said, there is room for more work on what is driving biased behavior (animus, statistical discrimination, or stereotypes), as this could help researchers and practitioners develop successful interventions.

There is some work on what works to reduce racial bias, and racial disparities more broadly, but the interventions that have been tested have mostly been unsuccessful. Many other interventions that are popular (implicit bias training, for instance) have yet to be rigorously evaluated. Experience in the criminal justice space and in other contexts shows us that many well-intentioned interventions will fail, and some will actively make the problem worse (see Doleac, 2020, for full discussion). It is crucial that practitioners and researchers work together to iterate upon new programs and policies until we find solutions that are effective and scalable.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. “Do Judges Vary in Their Treatment of Race?” *Journal of Legal Studies*, 41(2).

Agan, Amanda, and Sonja Starr. 2018. “Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment.” *Quarterly Journal of Economics*, 133(1): 191-235.

Albright, Alex. 2019. “If You Give a Judge a Risk Score: Evidence from Kentucky Bail Decisions.” Working paper.

Alesina, Alberto, and Eliana La Ferrara. 2014. “A Test of Racial Bias in Capital Sentencing.” *American Economic Review*, 104(11).

- Antonovics, Kate, and Brian G. Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *Review of Economics and Statistics*, 91(1).
- Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *Quarterly Journal of Economics*, 127(2).
- Anwar, Shamena, and H. Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review*, 96: 127-151.
- Arnold, David, Will Dobbie, and Crystal S. Yang. 2018. "Racial Bias in Bail Decisions." *Quarterly Journal of Economics*, 133(4).
- Arnold, David, Will Dobbie, and Peter Hull. 2020. "Measuring Racial Discrimination in Bail Decisions." Working paper.
- Ba, Bocar A. 2020. "Going the Extra Mile: The Cost of Complaint Filing, Accountability, and Law Enforcement Outcomes in Chicago." Working paper.
- Ba, Bocar, Dean Knox, Jonathan Mummolo, and Roman G. Rivera. 2021. "Diversity in Policing: The Role of Officer Race and Gender in Police-Civilian Interactions in Chicago." *Science*, 371(6530): 696-702.
- Becker, Gary S. 1957. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Carson, E. Ann. 2020. "Prisoners in 2019." Bureau of Justice Statistics, NCJ 255115.
- Canay, Ivan A., Magne Mogstad, and Jack Mountjoy. 2020. "On the Use of Outcome Tests for Detecting Bias in Decision Making." Working paper.
- Census, U.S. 2019. Quick Facts: Black or African-American, alone. Available at: <https://www.census.gov/quickfacts/fact/table/US/PST045219>
- Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires." *Quarterly Journal of Economics*, 131(3).
- Cohen, Alma, and Crystal S. Yang. 2019. "Judicial Politics and Sentencing Decisions." *American Economic Journal: Economic Policy*, 11(1): 160-191.
- Depew, Briggs, Ozkan Eren, and Naci Mocan. 2017. "Judges, Juveniles, and In-Group Bias." *Journal of Law and Economics*, 60(2).
- Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108 (2): 201-240.
- Doleac, Jennifer L. 2021. "A Review of Thomas Sowell's *Discrimination and Disparities*." *Journal of Economic Literature*, 59(2): 574-589.
- Doleac, Jennifer L. and Benjamin Hansen. 2020. "The unintended consequences of 'ban the box': Statistical discrimination and employment outcomes when criminal histories are hidden." *Journal of Labor Economics*, 38(2): 321-374.

Doleac, Jennifer and Sarah Lageson. 2020. "The problem with 'Clean Slate' policies: Could broader sealing of criminal records hurt more people than it helps?" Niskanen Center.

Eren, Ozean, and Naci Mocan. 2018. "Emotional Judges and Unlucky Juveniles." *American Economic Journal: Applied Economics*, 10(3).

FBI Uniform Crime Reports. 2019. "Table 43: Arrests." Available at: <https://ucr.fbi.gov/crime-in-the-u.s/2019/crime-in-the-u.s.-2019/topic-pages/tables/table-43>

Feigenberg, Benjamin, and Conrad Miller. 2021. "Racial Divisions and Criminal Justice: Evidence from Southern State Courts." *American Economic Journal: Economic Policy*, 13(2): 207-240.

Feigenberg, Benjamin, and Conrad Miller. 2020. "Racial Disparities in Motor Vehicle Searches Cannot be Justified by Efficiency." Working paper.

Gelbach, Jonah. 2021. "Testing Economic Models of Discrimination in Criminal Justice." Working paper.

Goncalves, Felipe, and Steven Mello. 2020. "A Few Bad Apples? Racial Bias in Policing." Accepted, *American Economic Review*.

Green, Tiffany L., and Nao Hagiwara. 2020. "The Problem with Implicit Bias Training." *Scientific American*.

Grogger, Jeffrey, and Greg Ridgeway. 2006. "Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness." *Journal of the American Statistical Association*, 101(475).

Harvey, Anna, and Taylor Mattia. 2021. "Reducing Racial Disparities in Crime Victimization: Evidence from Employment Discrimination Litigation." Working paper.

Heaton, Paul, Sandra Mayson, and Megan Stevenson. 2017. "The Downstream Criminal Justice Consequences of Pretrial Detention." *Stanford Law Review*, 69.

Heyes, Anthony, and Soodeh Saberian. 2019. "Temperature and Decisions: Evidence from 207,000 Court Cases." *American Economic Journal: Applied Economics*, 11(2).

Hoekstra, Mark, and CarlyWill Sloan. 2021. "Does Race Matter for Police Use of Force? Evidence from 911 Calls." Working paper.

Horrace, William C., and Shawn M. Rohlin. 2016. "How Dark is Dark? Bright Lights, Big City, Racial Profiling." *Review of Economics and Statistics*, 98(2).

Hull, Peter. 2021. "What Marginal Outcome Tests Can Tell Us About Racially Biased Decision-Making." Working paper.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. "Human Decisions and Machine Predictions." *Quarterly Journal of Economics*, 133(1): 237-293.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial bias in motor vehicle searches: Theory and evidence." *Journal of Political Economy*, 109(1).

Leasure, Peter, and Tara Martin. 2017. "Criminal Records and Housing: An Experimental Study." *Journal of Experimental Criminology*, 13(4): 527-535.

- Leasure, Peter, and Tia Stevens Andersen. 2016. "The Effectiveness of Certificates of Relief as Collateral Consequence Relief Mechanisms: An Experimental Study." *Yale Law and Policy Review Inter Alia*, November 7.
- Leslie, Emily and Nolan G. Pope. 2017. "The Unintended Impact of Pretrial Detention on Case Outcomes: Evidence from New York City Arraignments." *The Journal of Law and Economics*, 60 (3): 529–557.
- Linos, Elizabeth. 2018. "More Than Public Service: A Field Experiment on Job Advertisements and Diversity in Police." *Journal of Public Administration Research and Theory*, 28(1): 67-85.
- Luh, Elizabeth. 2020. "Not so Black and White: Uncovering Racial Bias from Systematically Misreported Trooper Reports." Working paper.
- McCrary, Justin. 2007. "The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police." *American Economic Review*, 97(1): 318-353.
- Miller, Amalia R., and Carmit Segal. 2019. "Do Female Officers Improve Law Enforcement Quality? Effects on Crime Reporting and Domestic Violence." *Review of Economic Studies*, 86(5): 2220-2247.
- Owens, Emily, David Weisburd, Karen L. Amendola, and Geoffrey P. Alpert. 2018. "Can You Build a Better Cop?" *Criminology & Public Policy*, 17(1): 41-87.
- Pager, Devah. 2003. "The Mark of a Criminal Record." *American Journal of Sociology*, 108(5).
- Park, Kyung. 2017a. "Do Judges Have Tastes for Discrimination? Evidence from Criminal Courts." *Review of Economics and Statistics*, 99(5): 810-823.
- Park, Kyung. 2017b. "The Impact of Judicial Elections in the Sentencing of Black Crime." *Journal of Human Resources*, 52: 998-1031.
- Piehl, Anne Morrison. 2009. "Preparing Prisoners for Employment: The Power of Small Rewards." Civic Report 57. New York: Center for Civic Innovation at the Manhattan Institute.
- Pfaff, John. 2017. *Locked In: The True Causes of Mass Incarceration and How to Achieve Real Reform*. Basic Books.
- Philippe, Arnaud, and Aurelie Ouss. "'No Hatred or Malice, Fear or Affection': Media and Sentencing." *Journal of Political Economy*, 126(5).
- Rehavi, M. Marit, and Sonja B. Starr. 2014. "Racial Disparity in Federal Criminal Sentences." *Journal of Political Economy*, 122(6).
- Rivera, Roman G., and Bocar A. Ba. 2019. "The Effect of Police Oversight on Crime and Allegations of Misconduct: Evidence from Chicago." Working paper.
- Roth, Jonathan, and Pedro H.C. Sant'Anna. 2021. "Efficient Estimation for Staggered Rollout Designs." Working paper.
- Schwartz, Joanna. 2020. "Allocating the Costs of Police Misconduct Litigation: Available Evidence and Research Agenda." Arnold Ventures.

Sherrard, Ryan. 2020. "Ban the Box' Policies and Criminal Recidivism." Working paper.

Sloan, CarlyWill. 2020. "Racial Bias by Prosecutors: Evidence from Random Assignment." Working paper.

Stevenson, Megan. 2018. "Distortion of Justice: How the Inability to Pay Bail Affects Case Outcomes." *Journal of Law, Economics, and Organization*, 34(4): 511-542.

Stevenson, Megan. 2019. "Assessing Risk Assessment in Action." *Minnesota Law Review*, 103.

Stevenson, Megan, and Jennifer L. Doleac. 2021. "Algorithmic Risk Assessment in the Hands of Humans." Working paper.

Tuttle, Cody. 2021. "Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums." Working paper.

Weisburst, Emily. 2017. "Whose Help is on the Way? The Importance of Individual Police Officers in Law Enforcement Outcomes." Working paper.

West, Jeremy. 2018. "Racial Bias in Police Investigations." Working paper.

Wood, George, Tom R. Tyler, Andrew V. Papachristos, Jonathan Roth, and Pedro H.C. Sant'Anna. 2020. "Revised findings for 'Procedural justice training reduces police use of force and complaints against officers'." Working paper.

Yang, Crystal S. 2015. "Free at Last? Judicial Discretion in Racial Disparities in Federal Sentencing." *Journal of Legal Studies*, 44(1).