Race, Context, and Judging on the Courts of Appeals: Race-Based Panel Effects in Death Penalty Cases

Jonathan P. Kastellec

Routledge
Taylor & Francis Group

Check for updates

Race, Context, and Judging on the Courts of Appeals: Race-Based Panel Effects in Death Penalty Cases

Jonathan P. Kastellec

Department of Politics, Princeton University, Princeton, NJ, USA

ABSTRACT

This paper examines how the identities of judges on multimember courts interact with case context to influence judicial decision making. Specifically, I leverage variation in panel composition and defendant race to examine race-based panel effects in death penalty cases on the Courts of Appeals. Using a dataset that accounts for several characteristics of a defendant and his crime, I find that the assignment of a black judge to an otherwise all-nonblack panel substantially increases the probability that the panel will grant relief to a defendant on death row—but only in cases where the defendant is black. The size of the increase is substantively large: conditional on the defendant being black, a three-judge panel with a single African-American judge is about 23 percentage points more likely to grant relief than an all-nonblack panel. These results have important implications for assessing the role of racial diversity on the federal courts and contribute to the empirical literature on the application of the death penalty in the United States.

Introduction

Judging is inherently a contextual act. While the task of legislators is to set broad policy to govern a range of situations, judges in a common law system—even when making legal policy that is broadly applicable—engage in case-by-case adjudication. The context of a case is multifaceted. First, each case contains distinct fact patterns, which can determine how judges apply or modify existing law. Second, different types of litigants, defendants, and attorneys may cause a case to be evaluated differently by judges and juries. Third, judges differ in their legal views, experience, and background, thereby creating differences in judicial evaluations (Harris and Sen 2019).

On trial courts, the fact that judges sit alone means that they decide cases without direct interactions with other judges. Trial judges with different ideologies and experiences may evaluate certain types of cases differently, but this evaluation will not be directly mediated by judicial colleagues.[1] Appellate courts, on the other hand, are multimember. Thus, judges with differing views and experiences may not only vary in their assessments of particular case contexts, but these differences will potentially influence the interactions among a group of judges who must produce a collective decision in a given case.

The relative importance of interaction between heterogeneous judges matters depends on the extent to which judicial diversity matters more broadly in appellate decision making. As a range of studies over the past two decades have made clear, the legal consequences of judicial diversity

---

CONTACT Jonathan P. Kastellec ✉ jkastell@princeton.edu 🖳 Department of Politics, Princeton University, Princeton, NJ, USA.
[1]In an interesting paper, Harris (2020) finds evidence that the proportion of black judges in a given courthouse affects the sentencing decisions of white judges. Such an effect, however, would likely not come through a case-specific intervention, but through a general diversity effect.

on multimember courts are indeed substantial. On the Courts of Appeals, the burgeoning literature on "panel effects" has documented that panels that are ideologically diverse are much more likely to make "moderate" decisions compared to ideologically homogenous ones (see e.g. Cross and Tiller 1998; Sunstein et al. 2006; Kastellec 2011). Similarly, in cases that implicate the interests of women or minorities, the addition of a single woman or minority judge, respectively, to an otherwise all-white or all-male three-judge panel significantly increases the likelihood of a liberal decision by the panel (Farhang and Wawro 2004; Cox and Miles 2008; Boyd, Epstein, and Martin 2010; Kastellec 2013).

From these studies, we know that the influence of judges on their colleagues on multimember courts is inherently contextual. Consider "partisan-based panel effects" (where the party of the appointing president is a proxy for ideology)—namely the finding that a judge is more likely to vote liberally with each additional Democratic colleague he sits with, and vice versa. These effects only operate in areas of the law that are sufficiently ideological such that liberal and conservative judges have latent disagreements over case outcomes (Sunstein et al. 2006). Next consider "characteristic-based panel effects"—that is, those related to the influence of racial or gender diversity. With respect to the latter, the addition of a woman to a panel increases the probability that men will vote for the plaintiff in sex discrimination cases, but makes no difference in cases not related to gender (Boyd, Epstein, and Martin 2010). Similarly, the two legal areas where race-based panel effects have been found (voting rights cases and affirmative action) are both tied to questions of race (Cox and Miles 2008; Kastellec 2013). In sum, judges tend to influence each other in areas of the law where we would expect such influence to occur.

Left unexplored, however, has been the possibility that panel effects on the Courts of Appeals are heterogeneous even *within* certain issue areas. In particular, in some areas, a case only carries ideological, racial, or gender dynamics under particular sets of case facts. We should thus expect panel effects in these areas to be *conditional* on salient features of a case that increase the likelihood that a judge will be able to influence her colleagues. This is particularly true with respect to *characteristic-based panel effects*, where the interaction between the characteristics of a judge and the facts of the case will be particularly clear.

In this paper, I examine how the identities of judges on multimember courts interact with case context to influence judicial decision making. Specifically, I study how the race of judges on three-judge panels interacts with the race of defendants in death penalty cases that reach the Courts of Appeals. The institution and history of capital punishment in the United States is inextricably tied to questions of race (Kennedy 1998, Ch. 9). Since the Supreme Court re-instituted the death penalty in 1976, a central debate surrounding the death penalty is whether the application of capital punishment is racially biased toward minorities—particularly African Americans, who are significantly overrepresented on death row relative to their share of the American population. At the same time, a large proportion of capital defendants are white. This means that the salience of race as an issue in a death penalty appeal is likely to vary with the race of the defendant.[2] Thus, with respect to variation in *cases*, a capital case involving a black defendant is likely to be contextually quite different from one involving nonblack defendants.

With respect to variation in *judges*, the racial nature of the debate over capital punishment means that, as an issue, the death penalty is potentially more salient for black judges, in particular. In turn, the nature of judicial interactions on a three-judge panel on the Courts of Appeals is likely to differ depending on whether or not the panel includes an African American judge. Due to their small numbers in the federal judiciary, black judges almost never comprise a majority of three-judge panels. Thus, the extent to which their inclusion on a three-judge panel is

---

[2]Another dimension on which the salience of race can vary is the race of the *victim* in the case (Baldus, Pulaski, and Woodworth 1983; Alesina and La Ferrara 2014). As I discuss below, the number of murders involving black victims that result in the death penalty is sufficiently small (particularly when the defendant is white) that making fine-grained statistical distinctions among different victim races is difficult.

consequential depends not on whether they vote differently from white judges, but whether they also influence their white colleagues (Kastellec 2013). Whereas other studies of panel effects have assumed constant effects within issue areas, the interaction between case and judicial context suggests that the likelihood of such influence will be conditional on the racial salience of each case. Because the issue of capital punishment is more explicitly racial when the defendant is black, we would only expect black judges to influence their nonblack colleagues in *cases where the defendant is black*.

I leverage the variation in both panel composition *and* defendant race to examine race-based panel effects in death penalty cases on the Courts of Appeals in which a defendant seeks relief from his death sentence in federal court. Using a dataset that accounts for several characteristics of a defendant and his crime, in addition to his race, I find that the quasi-random assignment[3] of a black judge to an otherwise all-nonblack panel substantially increases the probability that the panel will grant relief to a defendant on death row—but only in cases where the defendant is black. The size of the increase is substantively large: conditional on the defendant being black, a panel with a single African-American judge is 23 percentage points more likely to grant relief than a panel without any black judges. Finally, I show that this increase is not an artifact of the overall liberalism of the black judges on the Courts of Appeals—there is no parallel interactive effect between adding nonblack Democratic judges and the defendant being black. These results have important implications for assessing the role of racial diversity on the federal courts and contribute to the empirical literature on the application of the death penalty in the United States.

Race and judging in context

While the importance of case-specific context has long been recognized in the legal and political science literatures on judging, until recently judges on appellate courts were largely studied in isolation from their colleagues on a multimember panel. On the Courts of Appeals, in particular, the dominant mode of analyzing the relationship between judging and attributes like ideology, race, and gender was to simply compare the votes of judges from a particular group to those of another group, without recognizing the potential influences of collegiality on both individual and group decision making on a panel. As Farhang and Wawro (2004, 299) persuasively argue, an "accurate assessment of judges' attributes on panel decisions requires investigating whether and to what extent judges' attributes influence not only their own votes, but those of their panel colleagues."

The recent wave of panel effects studies makes clear the importance of such investigation. This is especially true with respect to characteristic-based panel effects, the study of which implicates the question of how well the increased *descriptive representation* of women and minorities on the federal bench translates into *substantive representation*—that is, better judicial outcomes for women and racial minorities compared to when nearly all federal judges were white men. Due to the relatively small numbers of women and minorities on the bench, only relatively rarely will they comprise a majority on a panel. Thus, the key quantity of interest is not individual differences in voting, but whether a judge's presence on a panel influences his or her colleagues' votes and thus results in a different case outcome (and a different majority opinion). As discussed above, a number of studies have documented such effects exactly in the issue areas where we would expect—that is, in cases where the interests of women or minorities are directly implicated.

At the same time, while the causal effect of ideological, racial, and gender diversity on panels is clear, the mechanisms underlying them are less well understood. Kastellec (2013) summarizes three possible mechanisms. The first is *deliberation*: judges with different life or professional experiences may offer different arguments to their colleagues, thereby changing the structure and

---

[3]As I discuss below in footnote 14, the assignment of judges to panels on the Courts of Appeals is not literally random, and hence I use the term quasi-random assignment throughout this paper.

outcome of a collegial decision by a three-judge panel (see e.g. Sunstein et al. 2006; Boyd, Epstein, and Martin 2010). The second is *votes*: the mere fact that women and black judges are more likely to vote liberally in the areas where characteristic-based panel effects have been studied may influence their colleagues to do so as well (see e.g. Fischman 2015). The third is *presence*: the mere characteristics of a judge who is different from the other two judges may cause the judges in the majority to evaluate a case differently, even independently of the votes cast in the case (see e.g. Kastellec 2013).

Adjudicating between these mechanisms is quite difficult. But the fact that characteristic-based panel effects have only been observed in areas where race and gender are salient is suggestive. For example, it is not the case that adding a woman to an otherwise all-male panel increases the probability that the male judges will vote liberally in cases involving, say, the Americans with Disabilities Act (Boyd, Epstein, and Martin 2010). Nor does adding a black judge to an otherwise all-nonblack panel in cases involving gender-based employment discrimination case increase the probability that the nonblack judges vote for the plaintiff (Farhang and Wawro 2004). Thus, it seems clear that the identity of the judge must interact directly with the issue-specific context of a case for this type of judicial influence to occur.

Recent studies in social psychology, economics, and political science on the role of race in jury decision making help illustrate how such a conditional effect might operate on the Courts of Appeals—at least, with respect to race. First, in a series of experiments, Sommers and Ellsworth (2000, 2001) have evaluated how variation in all of the following elements influence both individual and group-decision making (in the context of jury deliberations): the racial salience of a case, the race of a juror, the racial composition of a jury, and the race of the defendant. Specifically, the authors conducted mock trials in which they manipulated both the defendant's race and whether race is salient in the case.[4] When race is *not* salient, white jurors tend to evaluate black defendants more harshly. When race is *salient*, however, white jurors will evaluate black and white defendants equally. Black jurors, on the other hand, exhibited more leniency to black defendants, regardless of the racial saliency of the trial. The authors attribute the behavior of whites to a desire "to appear nonprejudiced when racial issues are salient" (Sommers and Ellsworth 2000, 1367).

Relatedly, in an observational setting, Anwar, Bayer, and Hjalmarsson (2012) study the relationship between the racial composition of jury pools (from which seated juries are drawn) and outcomes in jury trials. They find that juries formed from all-white jury pools are significantly more likely to convict black defendants than white defendants (conditional on case characteristics). However, the addition of even a single black member to the jury pool is sufficient to completely reduce the racial gap in conviction rates. Similar to the possible mechanisms for panel effects, one of the potential mechanisms here is that the addition of an African American member to a jury (which can only occur, obviously, if the jury pool contains at least one African American) influences how white jurors evaluate a criminal case involving a black defendant.[5] Similarly, in a study of death penalty trials, Bowers, Steiner, and Sandys (2001) find that as the

---

[4]For example, one of the experiments in Sommers and Ellsworth (2000, 1373) involved a man assaulting his girlfriend in a bar after she playfully made fun of him. In the race-salient treatment, the woman testified that the defendant had yelled, "You know better than to talk that way about a White [Black] man in front of his friends," before hitting her. In the treatment in which race was not salient, she testified that he had yelled, "You know better than to talk that way about a man in front of his friends."

[5]Anwar, Bayer, and Hjalmarsson (2012) also propose a second, indirect mechanism: the presence of African Americans in a jury pool may influence the calculus of the attorneys in a case of how best to use their peremptory challenges. "As a result, whenever an attorney uses a peremptory challenge to strike a black potential juror, she forgoes the possibility of excluding another potential juror with a similar ex ante likelihood of convicting. Put another way, even when black potential jurors are struck via peremptory challenges, they are essentially replaced on the jury by white jurors with similar attitudes toward the case" (Anwar, Bayer, and Hjalmarsson 2012, 1041).

ratio of whites to blacks on the jury increases, so does the likelihood of a black defendant being sentenced to death.

Finally, returning to the context of adjudication in appellate courts, a recent study by Grossman et al. (2016) of criminal appeals in Israel illustrates the potential effects of the interaction between panel composition and case type. The authors find that the quasi-random assignment of an Arab judge to a three-judge panel of Israel's appellate court increases the probability that the panel will rule in favor of an Arab defendant, compared to an otherwise all Jewish panel. No corresponding effect, however, for Jewish defendants is observed—their fate is independent of panel composition. Whether these findings can be extended to other courts, however, remains to be seen.[6]

Trial vs. appellate courts

Another important contextual distinction lies in the differences between trial courts and appellate courts—and how that distinction interacts with racial considerations. Because trial court judges sit alone, the question of whether minority and female judges tend to vote differently from white and male judges speaks directly to the extent of substantive representation on those courts. The evidence on this question is decidedly mixed. A range of studies have found no differences in voting behavior between minority and non-minority judges (Walker and Barrow 1985; Ashenfelter, Eisenberg, and Schwab 1995; Segal 2000), while others have found significant differences (Abrams, Bertrand, and Mullainathan 2012; Chew and Kelly 2006). In addition, a number of studies have evaluated how minority judges may condition their votes on different fact patterns in criminal cases—specifically, whether they tend to vote differently depending on the race of the defendant. Here again the evidence is mixed. For example, Welch, Combs, and Gruhl (1988) find that black judges sentence black defendants more leniently than white defendants. Similarly, Abrams, Bertrand, and Mullainathan (2012) find that judges tend to sentence African American defendants more harshly, on average, than white defendants; however, the gap in sentencing by defendant race is smaller among black judges than white judges. On the other hand, several studies have found that minority judges actually sentence minority defendants *more* harshly than white defendants (Uhlman 1978; Steffensmeier and Britt 2001; Spohn 1990; Fishman, Rattner, and Turjeman 2006).

When it comes to appellate courts, however, there are theoretical reasons to believe that black judges should be more likely to rule in favor of black defendants than white judges. The first relates to incentives. As Morin (2014) notes, most of these studies have focused on state courts, where judges must face reelection. Minority judges may have to worry about appearing too lenient when sentencing minority defendants. Federal judges, on the other hand, have life tenure and thus do not face such pressures. The second relates to case type. Whereas many criminal appeals are straightforward "easy cases" in which the law is clearly on the government's side, many involve claims of constitutional violations, such as cases involving search and seizure (Scherer 2004). And, indeed, the weight of the evidence supports the conclusion that minority judges are more likely to side with minority claimants. For example, Kastellec (2013) finds that African American judges almost always rule in favor of affirmative action when challenges to such policies are brought, while Morin (2014) finds that African American judges are more likely to vote in favor of black claimants (compared to other claimants) in employment discrimination cases. However, to my knowledge, no existing study has examined whether these individual differences translate to panel effects on the Courts of Appeals that are heterogeneous within a given area of the law.

---

[6]In a recent paper, Shahshahani and Liu (2017) find that that Jewish judges on the U.S. Courts of Appeals are significantly more likely than non-Jewish judges to favor claimants in religious liberties case, but they find no religious-based panel effects in these cases.

The death penalty context

Adjudication of capital claims provides a compelling forum for examining the intersection of panel effects and case characteristics. First, with respect to the U.S. Courts of Appeals, it is now clear that there exist sizable ideology-based panel effects in death penalty cases: both Fischman (2015) and Beim and Kastellec (2014) find that the likelihood of a pro-defendant vote by a given judge increases as she sits with a larger number of Democratic-appointed colleagues, and decreases as she sits with a larger number of Republican-appointed colleagues. The differences in the likelihood of a defendant receiving relief are substantial: panels composed of three Democratic appointees grant relief roughly 50% of the time, while panels composed of three Republican appointees grant relief only about 15% of the time (Beim and Kastellec 2014). Panels with at least one judge from each party, however, are more "moderate"—they grant relief between 20% and 30% of the time. Thus, it is clear that the quasi-random assignment of judges with different ideological backgrounds influences the outcomes of death penalty appeals.

No study, however, has examined the possibility of race-based panel effects in death penalty cases. As discussed above, such effects have only been documented in issue areas, like affirmative action and voting rights, where race is particularly salient and African American judges are likely to bring a different perspective to the adjudication of such cases. In those areas, where the salience of race is *constant*, the inclusion of a single black judge on an otherwise all-nonblack panel causes the nonblack judges to vote more liberally than they would otherwise, *ceteris paribus*.

How should we place capital punishment in the context of these findings? In theory, the administration of the death penalty could be orthogonal to questions of race. In practice, however, the institution of the death penalty in the United States has been inextricably tied to race. Since the death penalty was reinstated by the Supreme Court in 1976, perhaps the key debate about its implementation is whether racial bias in the legal system results in disproportionate numbers of minorities on death row—particularly African Americans, who are greatly overrepresented relative to their share of the American population. As of January 1st, 2020, of the roughly 2,600 inmates on death row, 42% were black, compared to a population rate of 14% (Fins 2020).

A plethora of studies both descriptively document these statistical disparities and test the harder question of whether these disparities result from systematic biases or from African Americans' greater propensity to commit crimes that result in death sentences (see e.g. Baldus, Pulaski, and Woodworth 1983; Baldus et al. 1997; Blume, Eisenberg, and Wells 2004; Radelet and Pierce 1991; Wolfgang and Riedel 1973). The most recent evidence that such biases exist comes from Alesina and La Ferrara (2014), who leverage both the race of the defendant *and* the race of the victim. It is well documented that not only are black defendants overrepresented on death row, but that black defendants whose victims are white are much more likely to receive a death sentence than black defendants whose victims are black (Radelet 1981; Blume, Eisenberg, and Wells 2004; Zeisel 1981; Fins 2020).[7] Alesina and La Ferrara (2014) find that appellate reversals of death sentences are more common when a minority defendant has killed a white victim, compared to a minority victim, a pattern consistent with trial judges and juries acting in a racially biased manner. Separately, Eberhardt et al. (2006) show that black defendants who are perceived as more "stereotypically black" based on their appearance are more likely to be sentenced to death, *ceteris paribus*.

Even if judges are not immersed in the statistical details of this debate, it seems likely that no judge can participate in the adjudication of capital punishment without being aware of the importance of race in the process—especially when the defendant is black. At the same time, whites comprise a roughly equal proportion (42%) of defendants on death row as blacks (Fins

---

[7]As Blume, Eisenberg, and Wells (2004) show, most capital crimes are intra-racial: about 85% of white victims are killed by white offenders, and over 90% of black victims are killed by black offenders. However, prosecutors are much more likely to seek the death penalty for black offenders who murder white victims, creating the imbalance in executions by race of the victim.
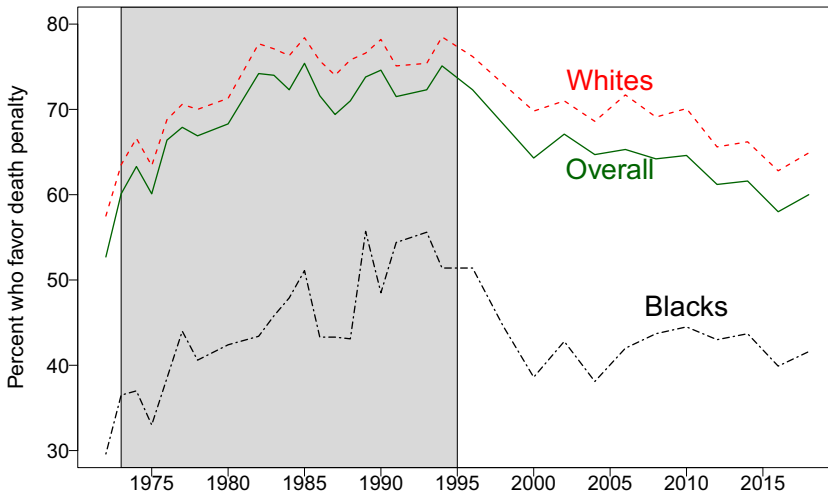
Figure 1. Public support for the death penalty, 1972 to 2018, as measured by the General Social Survey. The graph depicts responses to the question, "Do you favor or oppose the death penalty for persons convicted of murder?" (See https://tinyurl.com/ycx795vx for more details.) The line marked "overall" depicts support among Americans overall, while the other two lines break down opinion by whites and blacks. The shaded region depicts the scope of the Liebman dataset, 1973–1995.

2020); the difference, of course, is that whites are *underrepresented* relative to their share of the national population (61%). Combined with the fact that, as noted above, most white defendants have been convicted of killing white victims, this means that many cases will not implicate the disparities of African Americans on death row. In this sense, the context of a capital case involving a black defendant is likely to differ significantly from one involving a nonblack defendant.

Empirical predictions

How then should we expect race-based panel effects to operate in death penalty cases? If the death penalty were administered in a race-neutral manner, we would not expect the quasi-random assignment of African American judges to influence decision making of nonblack judges on appellate panels, since race would not be particularly salient, and thus black judges might approach a death penalty case no differently than a nonblack judge. Given how the death penalty is implemented in practice, however, the death penalty is likely to be more salient for black judges than nonblack judges (on average). When we look at mass opinion, for example, blacks have always been systematically less likely to support the death penalty than whites.

This can be seen in Figure 1, which depicts support for the death penalty as measured by the General Social Survey from 1972 to 2018. The figure shows that support for the death penalty has always been substantially higher among whites than blacks. Federal judges, of course, are not a random subset of the American population. However, as shown in Figure 1, the historical differences in opinion on the death penalty by race are so substantial that it seems reasonable to conjecture that black judges would tend to be more liberal on the issue than white judges, even after accounting for overall ideology.

Because a death penalty case becomes more racially salient when the defendant is black, we would expect the influence of a black judge on the votes of nonblack judges to be conditional on the race of the defendant. Under this theory, *the quasi-random assignment of a black judge to an otherwise all-nonblack panel should lead to more liberal voting, but only in cases where the defendant is black.* For cases with nonblack defendants, the prediction is less clear. Given the lack of racial salience in such cases, we might expect no differences across panel composition. Alternatively, Anwar, Bayer, and Hjalmarsson (2012) find that the addition of a black member to

a jury pool *increases* conviction rates for white defendants, to the point where the gap in the conviction rate for whites and blacks that exists when the jury pool is all white disappears. Applying a potential mechanism for this finding to the Courts of Appeals, the idea is that the quasi-random assignment of a black judge to a panel leads white judges to "be tougher" on white defendants than they would otherwise. Again, the prediction here is not clear cut. And, to foreshadow the results, there is no evidence for any such difference across panels with white defendants, in contrast to the significant evidence that panel composition matters for black defendants.

Data and results

To test the interaction between defendant race and panel composition in capital cases, I turn to the dataset of death penalty decisions compiled by James Liebman and his colleagues (see Liebman et al. 1999; Liebman, Fagan, and West 2000; Gelman et al. 2004). In an intensive data collection effort, these researchers collected information on the universe of capital appeals decided between 1973 and 1995. The dataset includes information on both appeals of death sentences heard by state and federal courts. (Alesina and La Ferrara (2014) use the Liebman et al. dataset—both state and federal court decisions—in their study of racial bias in capital sentencing.) The data on federal cases, which I use in this paper, contains information on all initial federal habeas appeals of a death sentence handed down by a state court. Under the doctrine of habeas corpus, a defendant can claim in federal court that his constitutional rights have been violated, and can seek redress from a federal court. This is a standard procedural move for capital defendants whose appeals have been denied in state appellate courts.

Using this dataset carries both advantages and disadvantages. The advantage is that Liebman and his colleagues painstakingly collected a wealth of information on each case, including numerous case facts related to the defendant (including race) and the circumstances of the crime.[8] To my knowledge, no other dataset of death penalty decisions by appellate courts contains this level of information. Accounting for these fact patterns is important, since different case profiles will lead to a higher or lower likelihood of relief for a defendant, regardless of the race of the defendant and the profile of judges hearing a case. Including these case facts should thus lead to more precise estimates of the effect of the interaction of panel composition and defendant race. In addition, the fact that all the cases in the data are habeas cases rules out any heterogeneity in decision making across different types of death penalty appeals.

The disadvantage of using the Liebman data is that it only extends to 1995. The cost of this temporal limitation is that there were relatively few African American judges on the federal bench in the time period covered by the data; in fact, only six unique black judges appear in the data.[9] The problem of low descriptive representation is common to studies of this sort; in their study of Israeli courts, for example, Grossman et al. (2016) have only seven Arab judges in their dataset. Note, however, that we are not interested in the voting behavior of these six judges *per se*, but in their influence on their nonblack colleagues. In the data, 47 unique nonblack judges sit with these six black judges—it is the behavior of these nonblack judges that determines the existence of race-based panel effects. In the appendix, I perform a cross-validation in which I replicate the main analysis presented below, each time dropping one black judge at a time. The results are robust to this procedure.

---

[8]The dataset is available from ICPSR (Fagan and Liebman 2006). The version archived there did not contain information on the race of the defendant and victim. I thank Valerie West for sending a version of the data containing this information.

[9]Between 1973 and 1995, the number of African American judges on the Courts of Appeals ranged from two to 12. In this period, 15 unique African American judges sat on the Courts of Appeals. As I note below, death penalty cases are concentrated in Southern states, so it is not surprising that the unique judges in the dataset are a subset of all black judges who served during this period.
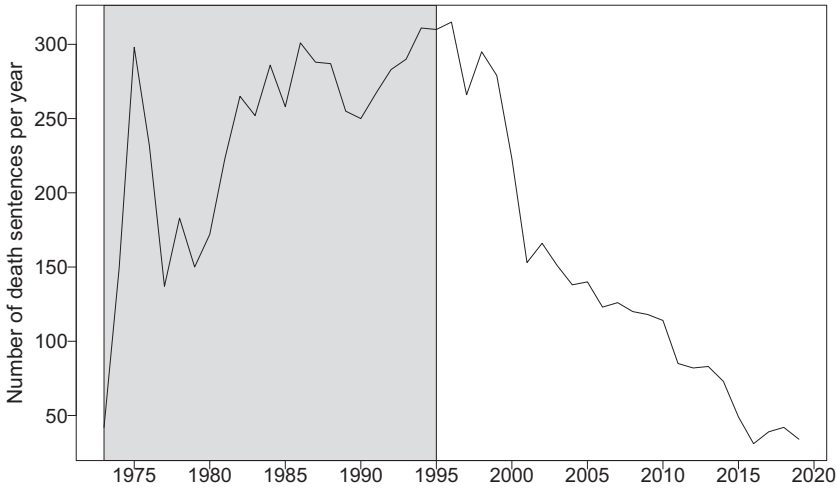
Figure 2. The number of death sentences per year in the United States, 1973–2018. The data was complied by the Death Penalty Information Center, based on annual reports published by the Bureau of Justice Statistics. (See tinyurl.com/ycr8yh8n for more details). The shaded region depicts the scope of the Liebman dataset, 1973–1995.

A second disadvantage is that the dataset ends just before the Antiterrorism and Effective Death Penalty Act of 1996 (AEDPA) was passed. This act sought to limit the power of federal judges to grant habeas relief. In another sense, however, this is an analytic advantage, as I analyze cases decided in an era where the capacity for judicial influence (and thus panel effects) is higher, due to the greater latitude federal judges possessed before the passage of AEDPA.

Despite these limitations, the period of 1973 to 1995 is an opportune time to study race-based panel effects in death penalty cases for other contextual reasons. Figure 2 depicts the number of death sentences in the United States between 1973 and 2018, with the shaded region covering the period of 1973 to 1995. The graph shows that the peak era of death sentences was in the 1980s and 1990s, with the number of capital sentences falling precipitously this century. While appellate review lags behind the imposition of sentences, such review was highly important during this earlier period—especially as federal judges had more discretion before AEDPA was passed.

In addition, if we return to Figure 1, we can see that the decades of the 1970s, 1980s, and 1990s represent the era in which the punitiveness of the American public reached its peak (Enns 2016). Support for the death penalty began to rise in the early 1970s and peaked in the middle of the 1990s, meaning that life-tenured federal judges were reviewing death sentences at a time when the public was highly supportive of capital punishment. In addition, just as the implementation of the death penalty in the United States inherently implicates race, so did many other facets of the criminal justice system in this period, including the War on Drugs, "three strike" laws, "truth in sentencing" laws, and mandatory minimum statutes (Alexander 2020). Thus, even while the time frame of the Liebman data is not ideal, it does cover an era in which the effects of racial diversity were likely to be particularly salient.

Description of variables

The Liebman habeas data contains information on cases heard at each level of the federal judicial hierarchy, including the district courts and the Supreme Court. Since my focus is on decision making on multimember courts with rotating panels, I select only cases heard by three-judge panels of the Courts of Appeals. This procedure results in a dataset of 515 cases decided by three-judge panels.

Dependent variable

The dependent variable I study is whether the panel granted habeas relief to the defendant, which results in the defendant's conviction or sentence being vacated, or both. *Relief* is coded 1 if the panel granted the defendant relief on any of his or her claims. (See the appendix for further details on the coding of this variable, as well as other variables.) The defendant was granted relief in 36% of cases; thus, only one of every three defendants received a positive outcome from the Courts of Appeals.

Judge and defendant race

For each case, I collected information on the three judges who heard the appeal, including their race. The first key variable for the analysis is *mixed panel*, which is coded 1 if a panel featured a black judge, and 0 if the panel was entirely nonblack (or "non-mixed"). Twelve percent of panels were mixed.[10] All of the mixed panels featured a single black judge; thus, on no panels did black judges comprise a panel majority.[11]

The next key variable is the race of the defendant. I code *black defendant* as 1 if the defendant was black and 0 if the defendant was nonblack. Consistent with the aggregate statistics discussed above, 44% of defendants were black, 52% were white, with the remainder Asian, Hispanic, or Native American.[12] Thus, a majority of cases feature nonblack defendants, but a significant proportion involve black defendants.[13]

Case characteristics

Following Gelman et al. (2004, 249–52), I use the following measures to account for different case characteristics and procedural histories that affect the underlying propensity for a defendant to be granted relief:

- *State evidentiary hearing*: Coded 1 if a state evidentiary hearing was held. Such a hearing improves the quality of factual inquiry, and should decrease the likelihood of habeas relief.
- *Out-of-state lawyer*: Coded 1 if the defendant was represented by an out-of-state lawyer. These lawyers are typically members of large firms who specialize in pro bono or special interest work, and thus their presence should increase the likelihood of relief.
- *Aggregative-mitigation index*: An index that accounts for the number of aggravating factors (such as murdering police officers) upon which jurors may lawfully consider as grounds for imposing a death sentence, minus the number of mitigating circumstances (such as intellectual disability) that the jurors may consider as a reason not to impose a death sentence.

---

[10]The only other minority group with representation on the Courts of Appeals in the period of study were Hispanics. (I use the term "Hispanic" to refer to persons in the United States who can trace their ancestry to the Spanish-speaking regions of Latin America and the Caribbean.) However, only four percent of panels featured a Hispanic judge, and only three percent of defendants were Hispanic. Accordingly, I pool Hispanic judges with white judges, and denote them collectively as nonblack judges.

[11]Of the 61 mixed panels in the data, a majority feature cases originating in four states: Florida (17), Georgia (14), Arkansas (11), and Missouri (7)—these states are all in either the 8th or 11th circuits. Each of these states has sizable African American populations. As of the 1990 census, the respective black population rate in these states was 14%, 27%, 16%, and 11% (McKinnon 2001). In practice, the modern day use of capital punishment has primarily been a Southern phenomenon (Death Penalty Information Center 2020), and Southern states tend to have larger black populations relative to the rest of the country.

[12]In the data, there are 18 Hispanic defendants, two Asian defendants, and three Native American defendants. Due to these small numbers, I pool these defendants with white defendants, denoting them collectively as nonblack defendants.

[13]Also roughly consistent with the aggregate data is the low proportion of black victims, as only 14% of victims in the data were black (this includes cases with multiple victims). Given the already small number of mixed panels, making fine-grained statistical distinctions across victim race is infeasible, and hence I focus solely on variation on defendant race. Based on the result in Alesina and La Ferrara (2014) discussed earlier, it seems plausible that the effect of mixed panels in black defendant cases might be heightened in cases where the victim is white, but this is purely a conjecture.

- *Offender-victim index*: An index of seven factors, separate from the formal aggravation and mitigation factors, that increase the perception of the seriousness of a capital offense. Examples of such factors include whether the defendant had a prior criminal record or was drunk at the time of the crime, and whether the victim was a woman or a police officer.
- *Federal evidentiary hearing*: Coded 1 if a federal evidentiary hearing was held prior to the final federal habeas review. Such a hearing should *increase* the probability of relief.
- *Number of claims*: The number of claims presented by the defendant in the habeas petition. Weaker defendants may employ a "kitchen sink" strategy of attempting to bring as many claims as possible, and so this variable should be negatively correlated with the likelihood of relief (Canes-Wrone, Clark, and Kelly 2014).
- *Year of sentencing*: The year in which the defendant was sentenced. Gelman et al. (2004) find that sentences imposed later were less likely to receive relief. I rescale this variable by subtracting 1973 (the first year of sentencing in the data) from it.

Finally, to account for ideological-based panel effects, for each judge, I code both the party of their appointing president and their ideology. For the latter I use the Giles, Hettinger, and Peppers (2001) measures of judicial ideology; I recode this variable such that higher values mean higher levels of liberalism. Based on these measure, I create the variable *panel liberalism*, which takes the mean of liberalism for all three judges. I also create a set of panel composition indicators: unified Republican (RRR), two Republicans and one Democrat (RRD), two Democrats and one Republican (DDR), and unified Democrat (DDD). The latter allows for possible non-linear changes in the propensity to grant relief across panel types (Beim and Kastellec 2014).

Estimation strategy and covariate balance

The research design I employed follows the general strategy of Grossman et al. (2016), who leverage quasi-random assignment of Arab and Israeli judges to cases in Israeli criminal courts. On the Courts of Appeals, the quasi-random assignment of judges to panels allows one to estimate the effect of panel composition (specifically, the difference between mixed and non-mixed panels)—but only conditional on the race of the defendants.[14] In other words, because of unobserved differences across black and nonblack defendants, one cannot simply ascribe any differences in rates of relief to racial motivations, even after including the above covariates. However, the quasi-random assignment of judges to panels reduces the hurdle of making a causal inference of the effect of mixed panels, conditional on defendant race. This is true because these unobserved differences that are correlated with defendant race are controlled for with this conditional assumption.

At the same time, given the time sweep of the data, the fact that assignment on the Courts of Appeals occurs *within* circuits, and that the distribution of black judges and black defendants may differ across circuits, it is important to assess how well balanced the key predictors are across the distribution of mixed panels. Table 1 depicts the mean rates of each predictor, including black defendant, across non-mixed and mixed panels. The second column from the right includes the *p*-value resulting from a difference of means test for a given predictor; the last column adjusts this *p*-value to account for the fact that the table includes multiple comparisons (the adjustments

---

[14]In reality, judges on the Courts of Appeals are not actually randomly assigned, as the procedures employed for panel assignment vary across circuits and allow for some discretion in panel selection, which mitigates against truly random selection (Hall 2010; Levy 2017). For instance, judges can trade places on panels in some circuits, and the original judges in a case that requires additional hearings may be selected for such subsequent hearings. In a recent paper, Chilton and Levy (2015) find deviations in a few circuits from a uniform distribution of panel assignments, where that distribution would be consistent with the underlying partisan distributions of those circuits. But there remains little evidence that assignment is ever systemically correlated with the desire to influence case outcomes, which is the key potential inferential concern.

Table 1. Balance statistics for panels with and without black judges. The *p*-value column depicts the results of a t-test comparing the means relevant variable across cases heard by mixed and non-mixed panels. The column "Corrected *p*-value" uses the method of Benjamini and Hochberg (1995) to adjust for multiple comparisons. After accounting for multiple comparisons, the only statistically significant difference is that panels with black judges feature more Democratic judges (and thus are more liberal overall), since all black judges in the data were appointed by Democratic presidents.

Variable	Mean, non-mixed panels	Mean, mixed panels	*p*-value	Corrected *p*-value
Black defendant	0.45	0.34	0.11	0.26
State evidentiary hearing	0.26	0.33	0.25	0.40
Out of state representation	0.32	0.38	0.41	0.52
Aggravation/Mitigation index	1.86	1.90	0.82	0.82
Offender victim index	2.07	2.02	0.74	0.82
Federal evidentiary hearing	0.20	0.31	0.05	0.14
Number of claims	4.48	4.87	0.40	0.52
Year sentenced	1980.00	1979.87	0.78	0.82
Number of Democrat judges on panel	1.33	2.07	0.00	0.00

use the method of Benjamini and Hochberg (1995); however, the results hold when any of the major adjustment procedures are used).

The table reveals generally good balance among all the predictors. Non-mixed panels were slightly more likely to hear cases with black defendants, but this difference is not statistically significant. In fact, the corrected *p*-values reveal that none of the differences are statistically significant, save for one: the average number of Democratic judges was much higher on mixed panels compared to non-mixed panels. This is not surprising, given that most black judges in the federal judiciary have been appointed by Democrats; and, in fact, this applies to all six black judges in the dataset.

The imbalance in partisanship across mixed and non-mixed panels necessitates ruling out the possibility that it is ideology *per se*, rather than the race of the judge, that is driving the changes in panel outcomes I observe. In fact, as I show below, there is no evidence for any interaction between panel ideology and defendant race. Relatedly, the fact that there are no black Republicans in the data creates complications in the interpretation of any race-based panel effects when controlling for ideology. This is because it is not possible to have a very conservative panel (i.e. one with three Republican appointees) that includes a black judge. I thus present models both with and without such controls—given the lack of any black Republicans, combined with the lack of any interaction between ideology and defendant race, the models without controls provide a cleaner estimate of the interaction between mixed panels and black defendants, *ceteris paribus*.

Estimation

To estimate the relationship between panel composition, defendant race and panel decision making, I use the following regression framework:

$$Pr(relief) = \text{logit}^{-1}(\beta_0 + \beta_1 * \text{MP} + \beta_2 * \text{BD} + \beta_3 \text{MP} \times \text{BD} + X\beta + \lambda C) \tag{1}$$

where *MP* denotes a mixed panel; *BD* denotes a black defendant, and $MP \times BD$ denotes the interaction between the two. X is the vector of case characteristics, while C denotes circuit fixed effects, which I include in most models.[15]

---

[15]In each model with circuit fixed effects, the 9th circuit serves as the omitted category. Because the death penalty has not been administered in many states, some circuits either do not appear in the data or have only very few cases. The 3rd, 6th, and 7th circuits all fall into this latter category; accordingly I pool cases from those circuits with those from the 9th circuit, and separately estimate fixed effects for the 4th, 5th, 8th, 10th, and 11th circuits.

Given the interaction term, interpreting the coefficients requires careful evaluation. For each panel-defendant combination, the likelihood of relief for each combination (conditional on the other covariates) can be assessed as follows:

- nonblack defendant, non-mixed panel: β_0
- nonblack defendant, mixed panel: $\beta_0 + \beta_1$
- Black defendant, non-mixed panel: $\beta_0 + \beta_2$
- Black defendant, mixed panel: $\beta_0 + \beta_1 + \beta_2 + \beta_3$

There are two sets of quantities of interest. First, for a given defendant race, one can calculate the effect of moving from a non-mixed to mixed panel. For nonblack defendants, this is simply β_1. For black defendants, it is $\beta_1 + \beta_3$. For the theoretical reasons discussed above, we would expect $\beta_1 + \beta_3 > 0$, meaning that the addition of a black judge to a case with a black defendant significantly increases the chance of relief.

In addition, the interaction term (β_3) itself is of substantive interest. There are two ways to interpret the interaction. First, looking from the direction of panel composition, a positive interaction would mean that switching from a nonblack to black defendant would have a larger effect among mixed panels. Second, looking from the direction of the race of the defendant, a positive interaction would mean that switching from a non-mixed to mixed panel has a larger effect among black defendants. Although one cannot empirically distinguish these symmetric interpretations (Brambor, Clark, and Golder 2006, 72), the latter is consistent with the relatively stronger theoretical expectations that mixed panels should have a larger effect on the propensity of relief for black defendants.

Finally, for nonblack defendants, from the effect of moving from a non-mixed to a mixed panel is given simply by β_1. As discussed above, the theoretical prediction for this effect is ambiguous.

Regression models and predicted probabilities

Table 2 presents four regression models; the dependent variable in each model is whether the panel granted relief or not. Model (1) allows *black defendant* and *mixed panel* to enter only as main effects, and not interactively. It also includes the case factors listed above, along with circuit fixed effects. Importantly, the coefficient on *mixed panel* is zero, meaning that, across all cases, there is no difference in the likelihood of voting for relief between panels that include a black judge and panels without a black judge, *ceteris paribus*. In addition, and unsurprisingly, the likelihood of relief increases with the overall liberalism of the panel.

What happens when we allow the effect of mixed panels to vary across defendant race? Model (2) in Table 2 includes only the interaction term *Mixed panel × black defendant*, along with the two constituent terms. I exclude any controls from these models. First, the main effect on *Mixed panel* is again statistically insignificant, meaning there is no shift in the likelihood of the panel voting for relief when the panel is mixed—conditional on the defendant being *white*. This result is consistent with race not being salient in cases with nonblack defendants.

Next, recall that the key test for the effect of mixed panels is whether the coefficients on the main effect on mixed panel and the interaction are greater than zero. The row ($\beta_1 + \beta_3$) in Table 2 reports the sum of these coefficients, while the row ($\beta_1 + \beta_3 = 0$) reports the p-value from an F-test of this null hypothesis. In Model (2), $\beta_1 + \beta_3 = 1.05$, and the p-value is .01 (one-tailed), meaning that the assignment of a single black judge on a three-judge panel significantly increases the likelihood of relief—for *black* defendants. (I return to a discussion of the interaction terms below.)

Model (3) retains the basic structure of Model 2, but also includes case characteristics and circuit fixed effects. The key results hold here as well. The main effect on *Mixed Panel* is statistically

Table 2. Logistic regressions of panel decisions to grant relief. Fixed effects for circuits are estimated in each model but are not displayed. Model (1) allows *black defendant* and *mixed panel* to enter only as main effects. Model (2) adds the interaction term *Mixed panel × black defendant*. Model (3) adds case characteristics; Model (4) also controls for *panel liberalism*.

	(1)	(2)	(3)	(4)
Intercept	2.52*	−0.60*	2.73*	2.70*
	(0.52)	(0.13)	(0.53)	(0.54)
Mixed panel	0.06	0.50	−0.09	−0.33
	(0.35)	(0.34)	(0.41)	(0.42)
Black defendant	−0.25	−0.16	−0.43	−0.40
	(0.22)	(0.20)	(0.23)	(0.24)
Mixed panel × black defendant		0.55	1.35*	1.16
		(0.58)	(0.67)	(0.68)
Panel liberalism	2.34*			2.25*
	(0.69)			(0.70)
State evidentiary hearing	−0.11		−0.07	−0.07
	(0.25)		(0.25)	(0.25)
Out of state representation	0.49*		0.59*	0.51*
	(0.23)		(0.23)	(0.23)
Aggravation/Mitigation index	−0.23*		−0.25*	−0.24*
	(0.08)		(0.08)	(0.08)
Offender victim index	−0.22*		−0.21*	−0.22*
	(0.09)		(0.09)	(0.09)
Federal evidentiary hearing	0.55*		0.49	0.52*
	(0.25)		(0.25)	(0.25)
Number of claims	−0.19*		−0.19*	−0.19*
	(0.04)		(0.04)	(0.04)
Year sentenced	−0.12*		−0.15*	−0.12*
	(0.04)		(0.03)	(0.04)
$B_1 + B_3$		1.05	1.26	.83
$B_1 + B_3 = 0$		.01	.01	.07
Circuit fixed effects?	Yes	No	Yes	Yes
N	515	515	515	515
log L	−221	−320	−225	−217

*denotes $p < .05$. The row $B_1 + B_3$ gives the sum of these coefficients, while the row $B_1 + B_3 = 0$ depicts the one-tailed p-value for the test of whether this sum is statistically different from zero.

insignificant, while $\beta_1 + \beta_3 = 1.26$; the hypothesis that $\beta_1 + \beta_3 = 0$ can be rejected at $p = .01$. Thus, even controlling for case characteristics, the assignment of a single black judge on a three-judge panel significantly increases the likelihood of relief for *black* defendants.[16]

Finally, Model (4) is the same as Model (2), but also includes *panel liberalism* as a control. As discussed above, the causal interpretation of the interaction between black judges on a panel and defendant race is murkier in these regressions, due to the fact that there are no black Republican judges in the data. The interaction and the main effect on mixed panel ($\beta_1 + \beta_3$) sum to .83, and the p-value on the null hypothesis $\beta_1 + \beta_3 = 0$ is .07 (one-tailed). Thus, the statistical significance of the difference across mixed panels for black defendants is somewhat weaker when panel liberalism is accounted for, but the basic pattern holds.

Finally, looking across the last three models in Table 2, the interaction term *Mixed panel × - black defendant* is statistically significant or close to it in Models (3) and (4); that is, the two models that include case characteristics and circuit fixed effects (the two-tailed p-values for the interaction term in Models (2), (3), and (4) are .34, .04, and .08, respectively.). This suggests that switching from a non-mixed to mixed panel has a larger effect among black defendants. Also worth noting is that the coefficient on *black defendant* is insignificant across all models—recall that the causal interpretation of this coefficient is not straightforward, but these results suggest

---

[16]In terms of the coefficients on the case characteristics themselves, each performs as it did in Gelman et al. (2004, 250, Table 6), except that whether a state evidentiary hearing was held does not significantly predict the likelihood of relief. One thing to note is that the two analyses differ slightly in that Gelman et al. use the defendant as the unit of analysis and ask whether relief was granted by any federal court, whereas I focus solely on decisions by three-judge panels of the Courts of Appeals.
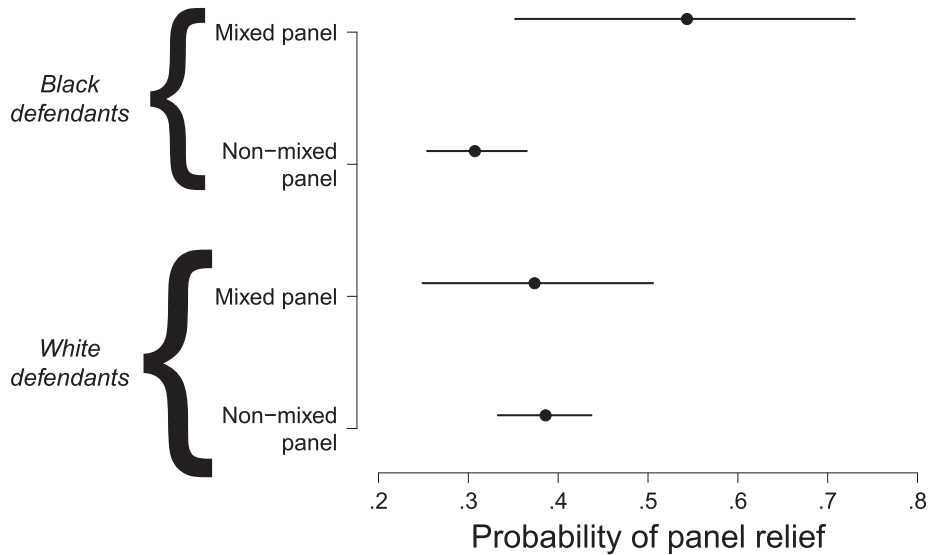
Figure 3. The substantive relationship between panel composition and defendant race. The points depict the average predicted probability of relief for each panel/defendant combination, based on the estimates in Model (3) in Table 2. Horizontal lines depict 95% confidence intervals (calculated via simulation).

that black defendants do not systematically fare worse in the Courts of Appeals than white defendants.

Predicted probabilities of panel relief

What is the substantive importance of the effect of racial diversity on panels for black defendants? Based on the estimates in Model (3), for each panel composition/defendant race combination, I calculate the "average predicted probability" of relief using the observed values in the data (Gelman and Pardoe 2007; Hanmer and Ozan Kalkan 2013). These probabilities are depicted in Figure 3. In addition, I use 1,000 simulations of the coefficients from Model (3) (based on variances and covariances of the coefficients) to generate confidence intervals on these predicted probabilities—these are depicted with the horizontal lines in Figure 3.

Beginning with black defendants, the probability of relief is .31 (95% confidence interval of [.25, .37]) if the defendant draws a non-mixed panel. The probability of relief rises to .54 [.35, .73] if the panel is mixed, a substantively very large difference of 23 percentage points. The 95% confidence interval of this difference is [.04, .43], and the difference is thus significantly larger than zero. With the caveat that there exists considerable uncertainty in the magnitude of the effect of mixed panels on cases with black defendants, it is useful to give this difference some context across issues: 23 percentage points is larger than the gender-based panel effects seen in sex discrimination cases (Farhang and Wawro 2004; Boyd, Epstein, and Martin 2010) and comparable to the race-based panel effects seen in affirmative action cases (Kastellec 2013).

Turning to nonblack defendants, the average probability of relief is nearly identical: .37 for mixed panels and .39 for non-mixed panels, with confidence intervals respectively of [.25, .51] and [.33, .44]. The lack of any difference is consistent with the ambiguous theoretical predictions regarding the effect of panel composition on nonblack defendants. Taken together, the results suggest the racial composition of a panel matters a great deal for black defendants, but does not matter for nonblack defendants.

Ruling out an alternative ideological story

As discussed above, black judges have traditionally been judicial liberals, and all the black judges in the dataset were appointed by Democratic presidents. Accordingly, it is important to consider the possibility that it is not race *per se* that induces the effects seen in cases with black defendants, but rather an ideology-based effect where more liberal judges influence their more conservative colleagues to vote more liberally when the defendant is black.

How might this occur? It is possible (and even likely) that capital cases with black defendants differ systematically from cases with nonblack defendants on dimensions beyond the observed covariates included in the regression analyses. These characteristics, which are observable to judges deciding cases but are unobservable to the researcher, may push cases involving a particular type of defendant closer toward a judge's "cutpoint," below which she would decide the death penalty is appropriate and above which she would decide it is not. Because habeas appeals of death penalty cases heard in federal courts often involve consideration of procedural issues, such as whether black jurors were properly excluded from the jury at the defendant's trial, black defendants may be more likely to be "marginal cases"—those on the threshold of receiving habeas or not.[17]

How would this unobserved shift in "marginality" interact with panel composition on the Courts of Appeals? Liberals, of course, are more likely to oppose the death penalty than conservatives; recall the huge difference in voting on death penalty cases between all-Republican and all-Democratic panels (Beim and Kastellec 2014). Thus, under this alternative story, the quasi-random assignment of *any* liberal judge, regardless of their race, may potentially move the panel closer toward a decision in favor of the black defendant. This would mean the race-based effect of adding a black judge would be a spurious effect of the fact that black judges tend to be liberals.

To assess this competing story, I present two regression models in Table 3 that allow for interactions between defendant race and ideological panel composition. These regressions can be thought of as placebo tests, where the placebo is ideological diversity and the treatment is racial diversity. Specifically, Model (1) in Table 3 interacts *Panel liberalism* with *black defendant*, while Model (2) uses partisan panel indicators (with RRR serving as the omitted category). In both models, to isolate the effect of ideology from any cross-effects of race, I drop all panels with black judges.

The results in both models are clear. In Model (1), the main effect of *Panel liberalism* persists, as the likelihood of relief is increasing in the liberalism of the panel. However, there is no interactive effect from the combination of increasing liberalism and black defendants—thus the effect of panel ideology is statistically indistinguishable across black and nonblack defendants. Similarly, in Model (2), the main effects on all three panel indicators are positive, meaning that panels with Democratic judges (of any number) are more likely to grant relief than panels with three Republican judges. However, all three interaction terms in Model (2) are insignificant (and, in fact, all three are in the "wrong" direction—i.e. negative).

To be sure, simply using party of the president as a proxy for ideology may mask important ideological differences between nonblack judges appointed by Democratic presidents and black judges appointed by Democratic presidents. However, if measurement error were the key driver of the difference between the main and placebo regressions, we would still expect to see *some* effect from an interaction of a black defendant and the participation of a nonblack Democratic judge on a panel. In the absence of *any* such effect, it seems clear that race is driving the panel effects when the defendant is black, not ideology.

---

[17]This argument is closely related to the theory developed by Park (2012), in which variation between levels of racial animus of judges and judge-specific legal standards is leveraged to test empirically for racial bias by trial court judges. In addition, the empirical results in Alesina and La Ferrara (2014) discussed above provide evidence that when black defendants kill white victims, their cases are more likely to be marginal.

Table 3. Assessing the alternative ideological story. Each model presents a logistic regression model of panel decisions to grant relief. Panels with black judges are omitted from the analyses. Fixed effects for circuits are estimated in each model but are not displayed.

	(1)	(2)
Intercept	2.72*	1.26
	(0.58)	(0.77)
State evidentiary hearing	0.07	0.09
	(0.28)	(0.28)
Out of state representation	0.64*	0.61*
	(0.25)	(0.25)
Aggravation/Mitigation index	−0.23*	−0.22*
	(0.09)	(0.09)
Offender victim index	−0.25*	−0.26*
	(0.10)	(0.10)
Federal evidentiary hearing	0.45	0.40
	(0.28)	(0.28)
Number of claims	−0.20*	−0.19*
	(0.04)	(0.04)
Year sentenced	−0.11*	−0.11*
	(0.04)	(0.04)
Panel liberalism	1.95*	
	(0.95)	
Black defendant	−0.37	0.28
	(0.31)	(0.76)
Panel liberalism × black defendant	0.39	
	(1.43)	
RRD panel		1.73*
		(0.62)
DDR panel		1.56*
		(0.64)
DDD panel		2.26*
		(0.79)
RRD panel × black defendant		−1.14
		(0.84)
DDR panel × black defendant		−0.31
		(0.86)
DDD panel × black defendant		−0.75
		(1.07)
N	454	454
log L	−183.96	−166.95

*indicates significance at $p < 0.05$.

Conclusion

This paper provides further evidence of the substantive consequences of judicial diversity on multimember courts. Leveraging the quasi-random assignment of black judges to three-judge panels, I find that the addition of a single black judge to an otherwise all-nonblack panel significantly increases the probability that a defendant gets relief from a death sentence—but only in cases where the defendant is black. No corresponding effect is seen for cases with nonblack defendants. Moreover, I find that this pattern is not simply an artifact of black judges being more liberal than nonblack judges. Taken together, the results provide additional evidence that racial diversity affects judicial interactions when a case implicates racial considerations.

With respect to criminal cases in general and death penalty cases specifically, the findings here (combined with the jury results seen in Anwar, Bayer, and Hjalmarsson (2012)) illustrate the effect of racial diversity throughout the legal process. Because one cannot easily judge the "correctness" of any given legal decision (e.g. whether a jury made the correct decision in sentencing a defendant to death), it is not possible to say that racial diversity leads to "better" outcomes. But it is clear that it leads to "different" outcomes. In addition, the incorporation of minority judges and jurors into the decision making process may enhance the overall legitimacy of the legal system, regardless of their effect on any single decision.

Moreover, while the issue of racial imbalances operates in all levels of the legal system, it is perhaps starkest with respect to capital punishment, given the severity and finality of such a sentence (if ultimately carried out). The small number of African American judges on the Courts of Appeals means that the effects documented here may not play out on a large scale in appellate judicial decision making. At the same time, the increasing presence of black judges on the federal bench certainly enhances the prospects for increased legitimacy in the appellate system among African American citizens.

In addition, the effects of racial diversity documented in the paper are likely to extend to state courts as well, which are much more active in both reviewing and reversing death sentences before many of them reach federal review (Gelman et al. 2004; Alesina and La Ferrara 2014). Testing for such effects is more difficult due both to data limitations and the fact that state supreme courts, which hear appeals of many death sentences, mainly sit *en banc* and not in rotating panels. Still, documenting the interaction between racial diversity on multimember courts and defendant race at the state level would enhance our understanding on the system-wide prevalence of the role of racial diversity in the judicial system, and thus would be a worthy pursuit for future research.

Finally, shifting back to panel effects more specifically, we still have much to learn about how judicial influence may be dependent on contextual factors. First, there are other areas of the law, such as employment discrimination, where the race of the parties varies, and thus so does the salience of race across cases—do the same patterns seen here hold in those areas? Second, does panel composition interact with case characteristics in other courts beyond the Courts of Appeals (Grossman et al. 2016)? Further study along these lines would increase our understanding of how case characteristics interact with the institution of multimember courts to influence judicial decision making.

Acknowledgments

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *Journal of Legal Studies* 41 (2):347–459.

Alesina, Alberto, and Eliana La Ferrara. 2014. "A Test of Racial Bias in Capital Sentencing." *The American Economic Review* 104 (11):3397–433.

Alexander, Michelle. 2020. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness.* New York: The New Press.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. "The Impact of Jury Race in Criminal Trials." *The Quarterly Journal of Economics* 127 (2):1017–55.

Ashenfelter, Orley, Theodore Eisenberg, and Stewart J. Schwab. 1995. "Politics and the Judiciary: The Influence of Judicial Background on Case Outcomes." *The Journal of Legal Studies* 24 (2):257–81.

Baldus, David C., Charles Pulaski, and George Woodworth. 1983. "Comparative Review of Death Sentences: An Empirical Study of the Georgia Experience." *Journal of Criminal Law and Criminology* 74 (3):661–753.

Baldus, David C., George Woodworth, David Zuckerman, and Neil Alan Weiner. 1997. "Racial Discrimination and the Death Penalty in the Post-Furman Era: An Empirical and Legal Overview with Recent Findings from Philadelphia." *Cornell Law Review* 83:1638.

Beim, Deborah, and Jonathan P. Kastellec. 2014. "The Interplay of Ideological Diversity, Dissents, and Discretionary Review in the Judicial Hierarchy: Evidence from Death Penalty Cases." *Journal of Politics* 76 (4): 1074–88.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57 (1):289–300.

Blume, John, Theodore Eisenberg, and Martin T. Wells. 2004. "Explaining Death Row's Population and Racial Composition." *Journal of Empirical Legal Studies* 1 (1):165–207.

Bowers, William J., Benjamin D. Steiner, and Marla Sandys. 2001. "Death Sentencing in Black and White: An Empirical Analysis of the Role of Jurors' Race and Jury Racial Composition." *University of Pennsylvania Journal of Constitutional Law* 3:171.

Boyd, Christina L., Lee Epstein, and Andrew Martin. 2010. "Untangling the Causal Effects of Sex on Judging." *American Journal of Political Science* 54 (2):389–411.

Brambor, T., W. R. Clark, and M. Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1):63–82.

Canes-Wrone, Brandice, Tom S. Clark, and Jason P. Kelly. 2014. "Judicial Selection and Death Penalty Decisions." *American Political Science Review* 108 (1):23–39.

Chew, Pat K., and Robert E. Kelly. 2006. "Unwrapping Racial Harassment Law." *Berkeley Journal of Employment and Labor Law* 27:49–110.

Chilton, Adam S., and Marin K. Levy. 2015. "Challenging the Randomness of Panel Assignment in the Federal Courts of Appeals." *Cornell Law Review* 101 (1):1–56.

Cox, Adam B., and Thomas J. Miles. 2008. "Judging the Voting Rights Act." *Columbia Law Review* 108 (1):1–54.

Cross, Frank B., and Emerson H. Tiller. 1998. "Judicial Partisanship and Obedience to Legal Doctrine: Whistleblowing on the Federal Courts of Appeals." *Yale Law Journal* 107 (7):2155–76.

Death Penalty Information Center. 2020. "Executions by State and Region Since 1976." https://deathpenaltyinfo.org/executions/executions-overview/number-of-executions-by-state-and-region-since-1976

Eberhardt, Jennifer L., Paul G. Davies, Valerie J. Purdie-Vaughns, and Sheri Lynn Johnson. 2006. "Looking death-worthy: perceived stereotypicality of Black defendants predicts capital-sentencing outcomes." *Psychological Science* 17 (5):383–6. doi: 10.1111/j.1467-9280.2006.01716.x.

Enns, Peter K. 2016. *Incarceration Nation*. Cambridge: Cambridge University Press.

Fagan, Jeffrey, and James Liebman. 2006. "Processing and Outcome of Death Penalty Appeals After Furman v. Georgia, 1973–1995: United States." doi: 10.3886/ICPSR03468.v1

Farhang, Sean, and Gregory Wawro. 2004. "Institutional Dynamics on the US Court of Appeals: Minority Representation under Panel Decision Making." *Journal of Law Economics & Organization* 20 (2):299–330.

Fins, Deborah. 2020. "Death Row U.S.A. Winter 2020." A Quarterly Report by the Criminal Justice Project of the NAACP Legal Defense and Educational Fund, Inc.

Fischman, Joshua B. 2015. "Interpreting Circuit Court Voting Patterns: A Social Interactions Framework." *Journal of Law, Economics, & Organization* 31 (4):808–42.

Fishman, Gideon, Arye Rattner, and Hagit Turjeman. 2006. "Sentencing Outcomes in a Multinational Society When Judges, Defendants and Victims Can Be Either Arabs or Jews." *European Journal of Criminology* 3 (1):69–84.

Gelman, Andrew, and Iain Pardoe. 2007. "Average Predictive Comparisons for Models with Nonlinearity, Interactions, and Variance Components." *Sociological Methodology* 37 (1):23–51.

Gelman, Andrew, James S. Liebman, Valerie West, and Alexander Kiss. 2004. "A Broken System: The Persistent Patterns of Reversals of Death Sentences in the United States." *Journal of Empirical Legal Studies* 1 (2):209–61.

Giles, Micheal W., Virginia A. Hettinger, and Todd Peppers. 2001. "Picking Federal Judges: A Note on Policy and Partisan Selection Agendas." *Political Research Quarterly* 54 (3):623–41.

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 (1):44–69.

Gryski, Gerald S, and Gary Zuk. 2008. "A Multi-User Data Base on the Attributes of U.S. Appeals Court Judges, 1801–2000." Accessed 15 August 2008. http://www.cas.sc.edu/poli/juri/auburndata.htm.

Gryski, Gerald S, Gary Zuk, and Sheldon Goldman. 2008. "A Multi-User Data Base on the Attributes of U.S. District Court Judges, 1789–2000." Accessed 15 August 2008. http://www.cas.sc.edu/poli/juri/auburndata.htm.

Hall, Matthew. 2010. "Randomness Reconsidered: Modeling Random Judicial Assignment in the U." *Journal of Empirical Legal Studies* 7 (3):574–89.

Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57 (1):263–77.

Harris, Allison P. 2020. "Can Racial Diversity Among Judges Affect Sentencing Outcomes?" Yale University working paper. https://www.allisonpharris.com/uploads/1/0/7/3/107342067/harris_divsent-4-2019.pdf.

Harris, Allison P., and Maya Sen. 2019. "Bias and Judging." *Annual Review of Political Science* 22 (1):241–59.

Kastellec, Jonathan P. 2011. "Panel Composition and Voting on the U.S. Courts of Appeals over Time." *Political Research Quarterly* 64 (2):377–91.

Kastellec, Jonathan P. 2013. "Racial Diversity and Judicial Influence on Appellate Courts." *American Journal of Political Science* 57 (1):167–83.

Kennedy, Randall. 1998. *Race, Crime, and the Law*. New York: Random House LLC.

Levy, Marin K. 2017. "Panel Assignment in the Federal Courts of Appeals." *Cornell Law Review* 103:65.

Liebman, James S, Jeffrey Fagan, and Valerie West. 2000. "A Broken System: Error Rates in Capital Cases, 1973–1995." *Columbia Law School, Public Law Research Paper*.

Liebman, James S., Jeffrey Fagan, Valerie West, and Jonathan Lloyd. 1999. "Capital Attrition: Error Rates in Capital Cases, 1973–1995." *Texas Law Review* 78:1839.

McKinnon, Jesse. 2001. "The Black Population: 2000." The U.S. Census Bureau. Accessed 15 May 2020. https://www.census.gov/prod/2001pubs/c2kbr01-5.pdf.

Morin, Jason L. 2014. "The Voting Behavior of Minority Judges in the US Courts of Appeals Does the Race of the Claimant Matter?" *American Politics Research* 42 (1):34–64.

Park, Kyung H. 2012. "Does Racial Animus Compromise Equal Protection? Evidence from Kansas Trial Court Judges?" University of Chicago Harris School of Public Policy Working Paper.

Radelet, Michael L. 1981. "Racial Characteristics and the Imposition of the Death Penalty." *American Sociological Review* 46 (6):918–27.

Radelet, Michael L., and Glenn L. Pierce. 1991. "Choosing Those Who Will Die: Race and the Death Penalty in Florida." *Florida Law Review* 43:1.

Scherer, Nancy. 2004. "Blacks on the Bench." *Political Science Quarterly* 119 (4):655–75.

Segal, Jennifer A. 2000. "Representative Decision Making on the Federal Bench: Clinton's District Court Appointees." *Political Research Quarterly* 53 (1):137–50.

Shahshahani, Sepehr, and Lawrence J. Liu. 2017. "Religion and Judging on the Federal Courts of Appeals." *Journal of Empirical Legal Studies* 14 (4):716–44.

Sommers, Samuel R., and Phoebe C. Ellsworth. 2000. "Race in the Courtroom: Perceptions of Guilt and Dispositional Attributions." *Personality and Social Psychology Bulletin* 26 (11):1367–79.

Sommers, Samuel R., and Phoebe C. Ellsworth. 2001. "White Juror Bias: An Investigation of Prejudice against Black Defendants in the American Courtroom." *Psychology, Public Policy, and Law* 7 (1):201–29.

Spohn, Cassia. 1990. "The Sentencing Decisions of Black and White Judges: Expected and Unexpected Similarities." *Law & Society Review* 24 (5):1197–216.

Steffensmeier, Darrell, and Chester L. Britt. 2001. "Judges' Race and Judicial Decision Making: Do Black Judges Sentence Differently?" *Social Science Quarterly* 82 (4):749–64.

Sunstein, Cass R, David Schkade, LisaM. Ellman, and Andres Sawicki. 2006. *Are Judges Political? an Empirical Analysis of the Federal Judiciary*. Washington, D.C.: Brookings Institution Press.

Uhlman, Thomas M. 1978. "Black Elite Decision Making: The Case of Trial Judges." *American Journal of Political Science* 22 (4):884–95.

Walker, Thomas G., and Deborah J. Barrow. 1985. "The Diversification of the Federal Bench: Policy and Process Ramifications." *The Journal of Politics* 47 (2):596–617.

Welch, Susan, Michael Combs, and John Gruhl. 1988. "Do Black Judges Make a Difference?" *American Journal of Political Science* 32 (1):126–36.

Wolfgang, Marvin E., and Marc Riedel. 1973. "Race, Judicial Discretion, and the Death Penalty." *The Annals of the American Academy of Political and Social Science* 407 (1):119–33.

Zeisel, Hans. 1981. "Race Bias in the Administration of the Death Penalty: The Florida Experience." *Harvard Law Review* 95 (2):456–68.

Appendix

Data and coding details

The Liebman data comprises the universe of "finally" reviewed habeas corpus death penalty cases from 1973 to 1995. Liebman, Fagan, and West (2000, 7) define the scope of the case selection as follows:

> For a federal habeas case to have been "finally reviewed" within the study period, all of the following events must have occurred in the case within that period: (1) a United States District Court must have denied habeas corpus relief from the capital judgment, thereby approving the judgment, or granted habeas relief from the capital judgment (either the conviction or sentence) on one or more grounds, (2) if an appeal was timely filed, a United States Court of Appeals must have approved or reversed the action of the District Court, and (3) if certiorari review was timely filed, the United States Supreme Court must have either (a) denied review or (b) granted review and affirmed or reversed the action of the Court of Appeals as of the first Monday in October 1995.

As discussed in the text, the unit of analysis in the Liebman et al. dataset is the defendant; for each defendant, information on his or her progress at every level of the judicial hierarchy is included. To study panel decision making, I reshaped the data such that every court decision is the unit of analysis, and retained only cases heard by three-judge panels of the Courts of Appeals.

Because in some cases the Liebman et al. coding of relief was based on an *en banc* or Supreme Court decision (which may have reversed a panel decision), I re-checked their coding of relief for every three-judge decision. The definition of relief (which, in practice, is analogous to reversal of a death sentence) can be found in Gelman et al. (2004, 218):

> Not all legal errors found by reviewing courts result in reversals. In most cases, reversal is permitted only if a death-row inmate can show (1) that a legal error committed in his or her case was "prejudicial" because there is a "reasonable probability that, but for the error, the outcome would have been different"; (2) that the error affected the verdict in some other identifiable way (i.e., was not "harmless"); or (3) that the error (e.g., bias against the defendant by the judge who presided at the trial) was "inherently prejudicial."

In a handful of cases I found coding errors, which I corrected.

For each case, I identified the three judges on the panel. Information on each judge's race, appointing president, party of the appointing president, home state, and year of appointment was taken from the appeals court judges attribute database (Gryski and Zuk 2008); for district court judges sitting by designation, the same information was taken from the district court judges attribute database (Gryski, Zuk, and Goldman 2008). In some cases, either a judge from the Federal Circuit or a non-Article III judge (for example, one from the U.S. Court of International Trade) sat on a three-judge panel. I used the biographical database of the Federal Judicial Center, available at http://www.fjc.gov/public/home.nsf/hisj, to identify the judge's race and party of his or her appointing president.

Robustness check: cross-validation of African American judges

As discussed in Section "Data and results", the fact that there are only six unique African American judges in the data raises the concern that the results may be idiosyncratic due to the characteristics of a given judge in the sample. As a robustness check, I performed a cross-validation where I replicated the analysis in Model (2) in Table 2 six times, each time dropping one of the African American judges from the analysis. These models are presented in Table A1. The key coefficients are very similar across the models. The p-value on the F-test of the null hypothesis that $(\beta_1 + \beta_3 = 0)$ is smaller than.05 in every model except Model (3), where it is.11 (one-tailed).

To confirm that the substantive results do not vary substantially across the cross-validated models, Figure A1 replicates Figure 3 in the paper, and depicts the average predicted probabilities by defendant race and panel composition for each of the six regression models presented in Table A1 (along with 95% confidence intervals). Figure A1 shows that while there is substantial variation in the average predicted probability of relief for black defendants when the panel is mixed, the basic pattern holds across the model: the likelihood of relief for a black defendant is greater when the panel contains at least one black judge, and the difference is of comparable magnitude to that seen in Table 3 in the paper. In addition, even when using Model (3)—which contains the noisiest estimates of $\beta_1 + \beta_3$—in 90% of simulations, the probability of relief when the panel is mixed (conditional on the defendant being black) is higher than the probability of relief when the panel is non-mixed. Thus, the key results are robust to excluding any one of the black judges in the data.
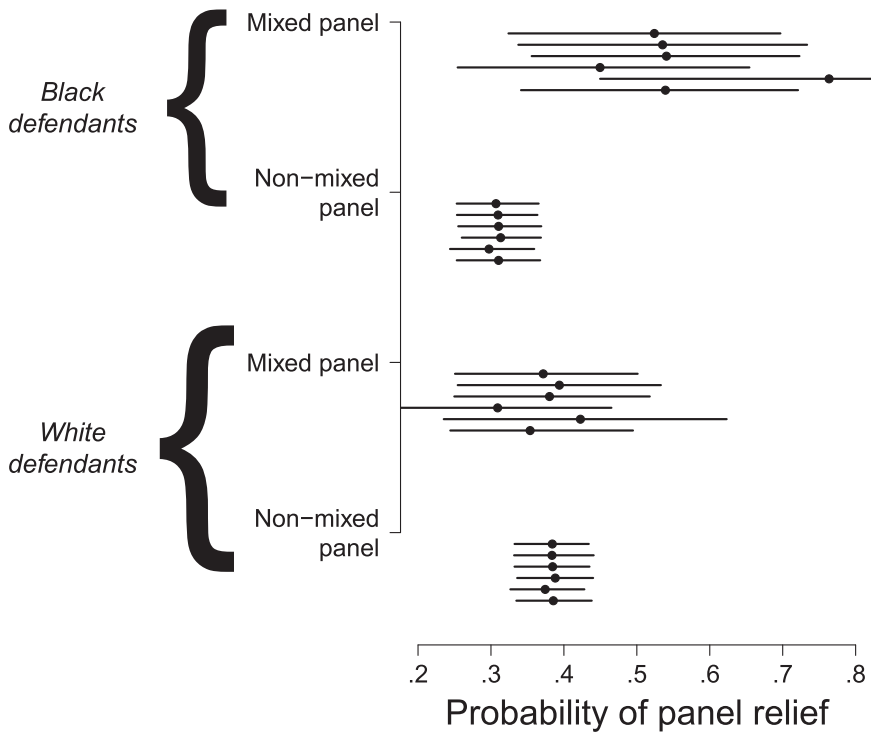
Figure A1. The substantive relationship between panel composition and defendant race, based on the results of each model in Table A1. The points depict the average predicted probability of relief for each panel/defendant combination, with 95% confidence intervals in brackets (calculated via simulation).

Table A1. Models of panel decisions to grant relief. In each model, one of the six unique black judges in the data is excluded. Fixed effects for circuits are estimated in each model but are not displayed.

	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	2.69	2.64	2.83	2.74	2.84	2.72
	(0.53)	(0.54)	(0.54)	(0.53)	(0.54)	(0.53)
Mixed panel	−0.16	0.27	−0.44	−0.02	0.06	−0.09
	(0.41)	(0.60)	(0.47)	(0.42)	(0.42)	(0.41)
Black defendant	−0.44	−0.45	−0.44	−0.43	−0.44	−0.43
	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)	(0.23)
Mixed panel × black defendant	1.41	2.34	1.21	1.29	1.23	1.27
	(0.68)	(1.03)	(0.76)	(0.68)	(0.68)	(0.68)
State evidentiary hearing	−0.06	0.00	−0.01	−0.09	−0.06	−0.06
	(0.25)	(0.26)	(0.26)	(0.25)	(0.25)	(0.25)
Out of state representation	0.60	0.73	0.61	0.57	0.60	0.59
	(0.23)	(0.24)	(0.23)	(0.23)	(0.23)	(0.23)
Aggravation/Mitigation index	−0.25	−0.25	−0.26	−0.25	−0.25	−0.25
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Offender victim index	−0.21	−0.23	−0.21	−0.20	−0.23	−0.21
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Federal evidentiary hearing	0.46	0.39	0.60	0.47	0.48	0.48
	(0.25)	(0.26)	(0.26)	(0.25)	(0.25)	(0.25)
Number of claims	−0.19	−0.19	−0.20	−0.19	−0.19	−0.19
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Year sentenced	−0.15	−0.15	−0.16	−0.15	−0.15	−0.15
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)
N	514	479	498	512	512	514
log L	−224.55	−204.09	−211.38	−224.30	−223.28	−224.84