Descriptive Representation and Judicial Outcomes in Multiethnic Societies

Guy Grossman University of Pennsylvania
Oren Gazal-Ayal University of Haifa
Samuel D. Pimentel University of Pennsylvania
Jeremy M. Weinstein Stanford University

The extent to which judicial outcomes depend on judges' identities is a central question in multiethnic societies. Past work on the impact of the racial composition of appellate courts has narrowly focused on civil rights cases in the United States. We expand this literature by testing for ethnicity-based panel effects in criminal appeals in Israel. Using randomness in the assignment of cases to panels, we find that appeal outcomes for Jewish defendants are independent of panels' ethnic composition. By contrast, panel composition is highly consequential for Arab defendants, who receive more lenient punishments when their case is heard by a panel that includes at least one Arab judge, compared to all-Jewish panels. The magnitude of these effects is sizable: a 14–20% reduction in incarceration and a 15–26% reduction in prison sentencing. These findings contribute to recent debates on the relationship between descriptive representation and substantive outcomes in judicial bodies.

A central challenge for multiethnic societies is how best to manage the interests of diverse groups. In these environments, politics often becomes an arena for competition among ethnic groups for access to power and the benefits that accrue to those that are victorious. Because the distinctions among ethnic groups are often seen as unalterable, exclusion from power can seem permanent. Notably, the consequences of exclusion are potentially manifold: inequalities in political voice and representation, inequities in the distribution of public resources, and disaffection with a state that is seen to represent some ethnic groups more than others.

Scholars have focused attention on the design of political institutions as a means of ensuring the representation of diverse groups while maintaining the legitimacy of state institutions (McEvoy and O'Leary 2013). Some have promoted the adoption of consensual institutions—such as proportional electoral systems and federalism—to promote the inclusion of minorities. Others have proposed policy remedies to address structural inequalities, including affirmative action, the use of ethnic quotas, and the construction of majority-minority electoral districts. Though still contested, an emerging body of evidence suggests that, in a wide variety of contexts, increases in the formal or descriptive representation of disadvantaged groups can translate into improvements in the substantive representation of their interests. For example, racial gerrymandering has been shown to generate more liberal representation in the southern United States (Shotts 2003), and gender quotas have contributed to changes in representation and public goods provision in India (Chattopadhyay and Duflo 2004).

Guy Grossman is Assistant Professor, Department of Political Science, University of Pennsylvania, 208 S. 37th Street, 225 Stiteler Hall, Philadelphia, PA 19104 (ggros@sas.upenn.edu). Oren Gazal-Ayal is Associate Professor, University of Haifa, 199 Abba Khushy Avenue, Haifa, Israel 3498838 (ogazal@univ.haifa.ac.il). Samuel D. Pimentel is Ph.D. Student, Statistics Department, University of Pennsylvania, 3730 Walnut Street, 400 Jon M. Huntsman Hall, Philadelphia, PA 19104 (spi@wharton.upenn.edu). Jeremy Weinstein is Associate Professor, Department of Political Science, Stanford University, Encina West, Room 100, Stanford, CA 94305 (jweinst@stanford.edu).

We thank Inbal Cabiri for excellent assistance in collecting the data used for this study, and Joshua Mitts for providing invaluable assistance in executing an automated text analysis of the court opinions. We also thank Larry Brown, Emil Pitkin, and Andreas Buja for helpful comments. Participants at the European Association of Law and Economics annual conference, the Conference on Empirical Legal Studies, and seminars at UNC Chapel Hill and Penn also provided valuable feedback. Research was conducted with generous support from Penn's Browne Center for International Politics and with partial government support under FA9550-11-C-0028 awarded by the Department of Defense, Army Research Office, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. Data for replication are available in the AJPS Data Archive on Dataverse (http://dvn.iq.harvard.edu/dvn/dv/ajps).

American Journal of Political Science, Vol. 00, No. 0, xxxx 2015, Pp. 1–26

© 2015 by the Midwest Political Science Association DOI: 10.1111/ajps.12187

For obvious reasons, much of this research has focused on institutions that are responsible for aggregating preferences, in which excluding minority groups is thought to be most consequential for policy choice (Lerner 2011). However, since the 1960s, there has been growing attention to the possibility that minorities receive unequal treatment in the judicial system, and to the political and social implications of inequalities in law enforcement. This line of inquiry is predicated on the assumption that unequal treatment by the courts can damage the legitimacy of the justice system, and thereby minorities' attitudes toward state institutions and their duties and responsibilities as law-abiding citizens. Indeed, evidence from a range of countries, as shown in Table 1, reveals large disparities in the incarceration rates of minority groups compared to the majority (a fact that in itself does not imply unequal treatment) and evidence of greater disaffection with the judicial system writ large. Over time, a discriminatory justice system can exacerbate social and economic inequalities and lead to a self-enforcing equilibrium in which antisocial behavior increases with expectations of racial or ethnic discrimination.

As in other institutional domains, one suggested remedy for reducing actual or perceived inequality in the justice system is to increase the diversity of the courts. This remedy might improve underrepresented groups' perceptions of the justice system if people value being judged by members of their own group, regardless of the outcome (Ifill 2000). The expectation that increasing the representation of minorities will result in changes in substantive outcomes is, however, more contested in the judicial sphere. While minority judges may bring a different judicial perspective to the bench, judges are expected to administer the law in an impartial manner. For a professional body that allegedly interprets and applies the law objectively, the substantive impact of changes in the ethnic or racial balance of judicial representation is far from obvious. Studies that explore whether judicial outcomes depend on judges' background characteristics have mostly focused on how judges' identities shape decision making on issues of civil rights—where the claim that minorities have a different judicial perspective is quite compelling. By contrast, the evidence that descriptive representation leads to different substantive outcomes in criminal sentencing is much less clear.

In this article, we exploit a natural experiment in Israel in which the diversity of panels in appellate courts is quasi-randomly assigned, such that criminal defendants sometimes face a panel composed entirely of judges from the majority group (i.e., Jewish) while at other times panels include a minority (i.e., Arab) judge. We collected information on all panel rulings in sentencing appeals in criminal cases in district (appellate) courts in Israel between 2007 and 2011.[1] Specifically, we focus on three district courts—Tel Aviv, Nazareth, and Jerusalem—that had at least one Arab judge hearing criminal appeals during that period. Since judges are assigned by the court clerk to hear criminal appeals on specific dates (independently of what is on the court's docket), the assignment of judges to panels and panels to cases is "as good as random." As we demonstrate below, the resulting ethnic composition of multi-judge panels is orthogonal to case characteristics.

Our identification strategy allows us to estimate the causal effect of ethnicity-based panel composition, but only *conditional on the defendant's ethnicity*. Though we are able to identify the causal effect of mixed panels *separately* for Jewish and Arab defendants, we do not estimate a gap in punishment *between* Jewish and Arab defendants.[2] This is because there are likely systematic differences between Arab and Jewish defendants, at all stages of the criminal justice system, that are unobserved by the researchers and therefore cannot be controlled for (Abrams, Bertrand, and Mullainathan 2012).

Our empirical analysis advances in two steps. First, we examine the likelihood that an appellate court will overrule a decision made by the lower magistrate court as a function of the defendant's ethnicity and the identity of the appellant (i.e., the defendant or the prosecution). We find robust evidence that mixed panels (i.e., those that include at least one Arab judge) are significantly more likely to accept the sentencing appeals of Arab defendants and reduce the sentence handed down by the magistrate court. By contrast, the likelihood that Jewish defendants' appeals will be accepted by the district court is independent of the panel's ethnic composition. Moreover, we find that mixed panels are less likely to accept the prosecution's appeal to increase the sentence handed down by the

[1]This study focuses on appeals against *sentence severity* rather than against *convictions* for two reasons. First, judges exercise more discretion in sentencing than in determining guilt. In determining guilt, judges seek to identify "juridical truth"—whether there are reasons to believe beyond a reasonable doubt that the defendant committed the offense. By contrast, judges exercise "stronger" discretion in sentencing since there are no right or wrong answers. Since we are interested in the effect of ethnic diversity on judicial outcomes, it makes sense to examine cases in which discretion is high. Second, successful appeals against the verdict are uncommon. By contrast, sentencing appeals are not infrequent, which leads to a greater degree of variation necessary to identify the effect of panels' ethnic composition, if such an effect exists.

[2]Our findings, however, provide suggestive evidence that such a gap exists. Also note that our data do not allow us to examine whether mixed panels reach "better" outcomes from a social welfare perspective.

Table 1 Incarceration Rates and Support of the Justice System

Country	Ethnic/Racial Group	Population Share	Incarceration Rate (per 100,000)	Support of Justice System
Australia	Non-indigenous	97	130	54
	Indigenous	3	1868	46
Israel	Jews	75	86	56
	Arabs	21	199	49
South Africa	Blacks	79	332	71
	Colored	9	650	54
United States	Whites	80	340	61
	Blacks	13	2362	41

Note: Data on support of the justice system for the United States, South Africa, and Australia are derived from the World Value Survey. Incarceration rates data come from various sources: United States (United States Census Bureau, see http://www.census.gov/popest/data/historical/2000s/vintage_2006/; Bureau of Justice statistical tables, see http://www.bjs.gov/content/pub/pdf/pim09st.pdf), Australia (Australian Bureau of Statistics, see http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/2075.0Main%20Features42011?opendocument&tabname=Summary&prodno=2075.0&issue=2011&num=&view=), South Africa (Department of Correction Services, see http://www.abs.gov.au/ausstats/abs@.nsf/0/875C813AF74635EBCA25795F000DB4EF?opendocument), and Israel (Korn 2003).

magistrate court, but again only in cases involving Arab defendants. Whereas the prosecution "wins" 83% of its appeals in all-Jewish panels involving Arab defendants, the figure drops to about 63% when the panel includes at least one Arab judge.

In the second stage, we quantify the implications of the above findings. Focusing on incarceration rates, we find that Arab defendants are significantly less likely to be incarcerated by a mixed panel than an all-Jewish panel. The magnitude of this difference is about 17% using our preferred model. This finding is robust to various model specifications and to the inclusion of a large set of defendant, case, court, and judge controls. We do not, however, find evidence of panel effects on incarceration for Jewish defendants. Moving to the length of the prison term, we find that for both Jewish and Arab defendants, mixed panels result in reduced prison terms compared to all-Jewish panels. The magnitude of this effect is insignificant for Jewish defendants. For Arab defendants, the reduction in prison sentence is larger and highly significant: roughly 2.7–4.7 months, or a 15–26% reduction in the length of the average prison sentence. These results are particularly striking, given the lower incarceration rates of Arab defendants by mixed panels, which essentially means that mixed panels sentence Arab defendants to imprisonment in *more severe* cases.

This article contributes to the study of the determinants of judicial decision making. Specifically, the article joins calls by Epstein and Knight (2013) and others to go beyond the long-held notion among political scientists that judges are policy oriented—that is, that rulings are first and foremost influenced by judges' *political views*. Whereas the article certainly is situated within a legal tradition that rejects the idea that judges are apolitical, nonpartisan, value-free umpires (Mendelson 1964), it focuses on determinants of decision making that cannot be easily reduced to ideological or partisan ends. Instead, we provide further evidence of the importance of identity consideration and institutional norms, which too often have been overlooked by judicial politics scholars (Epstein and Knight 2013).

The findings of this article also contribute to the literature on the implications of judicial diversity. We build on the results of past studies that have shown that panel composition (whether examined by gender or race) has a significant causal effect on judicial outcomes. Past studies, however, have examined race-based panel effects only in civil rights cases. To the best of our knowledge, ours is the first article to demonstrate the impact of judicial diversity on the outcomes of criminal appeals, which do not necessarily have a specific ethnic content.[3] An additional limitation of past panel effects studies is that they all use data from the U.S. justice system. However, to test whether theories of panel effects are generalizable, we ought to move beyond a limited geographic focus. By situating our study in Israel, we expand the geographic and legal context in which panel composition seems to matter.

Finally, the article makes a contribution to the theoretical and policy debate on the relationship between descriptive and substantive representation in the judiciary. In Israel, for example, there is mounting pressure

[3]Eisenberg, Fisher, and Rosen-Zvi (2013) also study panel effects in criminal appellate courts, focusing on whether judges' votes are affected by the attitudes and preferences of the presiding justice.

to make the judiciary more representative of minority and other disadvantaged groups. To date, policy makers and the professional bureaucracy at the Ministry of Justice object to such calls, arguing that the substantive representation of minorities is best achieved not via greater representation but through an independent judiciary. While the evidence in support of descriptive representation is increasingly strong for political institutions, its value in the judicial sphere remains contested. The evidence in this article—that changes in the ethnic composition of the judiciary matter immensely for sentencing outcomes—is important as policy makers grapple with pressures to increase the diversity of professional state institutions in multiethnic societies.

The Impact of Judicial Diversity

This article contributes to a growing literature on judicial diversity, which explores the conditions under which minority defendants and claimants fare better (or worse) when members of their own group hear their cases. Most scholarly work focuses on single-judge rulings, for which the evidence of an ingroup preference is quite mixed. A new body of work focuses on multijudge settings, but the emphasis is primarily on how racial and gender diversity shapes decisions on issues that are highly salient to the disadvantaged groups, such as civil rights cases, which have a high expectation that descriptive representation will lead to different substantive outcomes. Research on whether descriptive representation changes substantive outcomes in multi-judge criminal appeals is still lacking.

Single-Judge Rulings

When court judges are solitary decision makers, the mapping from descriptive representation to substantive outcomes is straightforward, and it boils down to whether female and minority judges rule differently than male and majority-group judges in similar cases (Kastellec 2013). This is, therefore, a natural starting point for examining the effect of judges' backgrounds on judicial outcomes (Nagel 1962). Reviewing a large body of work, the evidence suggests that whether minority judges favor ingroup members crucially depends on specific characteristics of the legal procedure.

The literature focuses on two factors that impact the extent of ingroup favoritism in single-judge rulings and *pull rulings in opposite directions.* On the one hand, identity considerations pull racial minorities and female judges to favor ingroup members, especially in areas of the law that are highly salient to members of the disadvantaged group (Farhang and Wawro 2004). This effect might be due to personal preferences ("taste" mechanism) or to different life experiences that make minority and female judges more sympathetic to claims made by members of their own group (Martin 1990). On the other hand, judges operate in institutional and organizational settings that discourage deviations from institutional norms, such as judicial impartiality, respect for precedent, and dissent aversion (Epstein and Knight 2013). These norms are designed to constrain judges from taking advantage of the discretion afforded them. The evidence suggests that how judges balance these opposing factors depends on the nature of the judicial procedure—whether it involves (a) fast, low-stakes procedural decisions; (b) civil rights cases, which are particularly salient to members of disadvantaged groups; or (c) decisions on criminal sentencing. These legal settings differ from one another in a number of ways, including the stakes of the decision, the speed with which decisions are made, the salience for minority groups, and the extent of discretion permitted to judges as they fashion their rulings.

There is robust evidence that in single-judge rulings that entail relatively low stakes and fast decision making, minority judges are significantly more likely to rule in favor of ingroup members. Building on social identity theories (e.g., Tajfel and Turner 1979), it has been argued that fast, nondeliberative decision making legal settings are especially subject to implicit biases that produce ingroup favoritism.[4] Ingroup favoritism by minority judges has been identified, for example, in small claims courts (Shayo and Zussman 2011), detention (Gazal-Ayal and Sulitzeanu-Kenan 2010), and bail-setting decisions (Ayres and Waldfogel 1994), as well as in quasi-juridical rulings of officials in basketball (Price and Wolfers 2010) and baseball leagues (Parsons et al. 2011). We argue that the nondeliberative and low-stakes nature of these procedures creates an environment in which judges are more likely to follow their personal (explicit or implicit) preferences while discounting, deliberately or unconsciously, institutional constraints on their discretion.

Work in social psychology suggests that decision makers can (and often do) overcome implicit biases in cases that require careful deliberation (Rachlinski et al. 2009). There is, however, some evidence that minority and female judges continue to exhibit ingroup favoritism, even in high-stakes deliberative procedures on issues that are highly salient to the disadvantaged group. For example, even though across a large number of legal settings most studies do not find differences in ruling patterns

[4] Implicit bias is defined as an individual's unconscious positive or negative mental attitude toward a person, thing, or group.

between white and black judges (Ashenfelter, Eisenberg, and Schwab 1995), there is evidence of ingroup favoritism among black judges in racial harassment cases (Chew and Kelley 2006). Similarly, though Boyd, Epstein, and Martin (2010) do not find differences in gender-based voting patterns across 13 areas of the law, they do identify significant differences in how male and female judges rule in sex discrimination cases. Findings from studies on state supreme courts also show that female justices are more likely than their male counterparts to support plaintiffs in sex discrimination cases (Allen and Wall 1993; Gryski, Main, and Dixon 1986). Here, despite the high stakes and careful deliberation that decisions entail, the salience of the issue to disadvantaged groups may drive judges to take account of their implicit or explicit biases when making a ruling.

Finally, evidence of ingroup favoritism is least robust in the literature on criminal sentencing decisions. For example, scholars generally do not find differences between male and female judges at the state trial court level (Gruhl, Spohn, and Welch 1981). Turning to racial differences, several studies—including Welch, Combs, and Gruhl (1988), Johnson (2006), and Holmes, Hosch, Daudistel, Perez and Graves (1993)—find evidence of ingroup preference among racial minority judges. However, Uhlman (1978) finds no differences in rulings between white and black judges, and Fishman, Rattner, and Turjeman (2006) find that in Israel, Arab judges were harsher than Jewish judges toward Arab defendants.[5] Spohn (1990) also finds that black judges are slightly harsher than white judges, but she does not find any ingroup or outgroup effects. Reviewing these and other studies, Spohn (2009, 121) concludes that "the evidence regarding the degree to which black and Hispanic judges impose more racially equitable sentences is . . . mixed." Schanzenbach (2005, 59) similarly concludes that in criminal sentencing, "the literature on judging has not produced consistent findings regarding the impact of a judge's race on his or her decisions."

There are a number of reasons for the absence of a clear-cut ingroup bias in criminal cases. First, as mentioned, the fact that criminal cases involve more serious deliberation than procedural rulings may help judges overcome their implicit biases. Second, some features of criminal sentencing privilege the impact of institutional factors over identity considerations. For example, minority judges may be especially reluctant to be perceived as favoring their group in criminal cases, to which the public and media are often more attuned. The strong professional pressure to rule impartially in the criminal court system—compared to civil rights cases, which are commonly politicized—thus helps neutralize race and gender effects (Hilbink 2007). Third, minority judges may have internalized the prevailing view of members of the majority, who often perceive minority defendants as more dangerous than majority-group defendants (Fishman, Rattner, and Turjeman 2006). Alternatively, minority judges may see harsh sentencing against ingroup members as consistent with their group interest. For example, there is some evidence of harsher sentencing of black defendants by black judges in drug-related crimes (Steffensmeier and Britt 2001).

The above factors may encourage judges to deemphasize identity considerations in criminal sentencing compared to both low-stakes and nondeliberative procedures and civil rights violations. Consistent with this argument, in one of the most methodologically sophisticated studies to date, Abrams, Bertrand, and Mullainathan (2012) find that between-judge variation in attitudes toward sentencing is significantly more important than judges' background characteristics in explaining sentencing variability. In sum, the evidence suggests that the type of legal setting impacts the way in which judges trade off identity considerations against institutional norms and organizational pressures. We have argued that single-judge criminal sentencing is an ideal-typical setting in which identity considerations (which, on average, lead to ingroup favoritism) have relatively weaker force compared to institutional considerations, leading judges to greater conformity with institutional norms.

Though a framework that highlights how legal procedures facilitate or constrain judicial bias can be generalized to diverse settings, there is important cross-country (and cross-state) variation in the relative strength of the organizational structures and norms that are designed to limit judges' discretion. For example, the discretion of judges is likely influenced by the security of tenure as well as the extent to which hiring and promotion decisions are allowed to be politicized. Politicized hiring and promotions systems, which are common in the United States, give judges more leeway to decide according to their ideological convictions and personal preferences. By contrast, in settings in which promotions are largely depoliticized, such as in Israel, judges may be relatively more constrained in their ability to allow personal preferences to impact their rulings.

Panel Settings

In recent years, there has been increasing interest in complementing the analysis of single-judge decision

[5]This finding may be spurious, as it relies on a doubtful coding scheme in which cases decided by panels were treated as single-judge rulings, and the judge ethnicity in panel cases was determined by the nationality of the majority. This means that panels with two Jewish judges and one Arab judge were coded as Jewish, whereas we code this as a mixed panel.

making with an examination of the relationship between descriptive representation and substantive outcomes in multijudge settings. This shift has been justified by the fact that standard approaches that look narrowly at how judges' ascriptive characteristics affect their rulings are insufficient to capture the dynamics of multi-judge settings (Farhang and Wawro 2004). This is because the judges' identities (Peresie 2005) and votes (Fischman n.d.) likely impact the decision making of other judges on a panel, including by potentially changing the strategic dynamic and how judges trade off their own personal preferences and institutional and organizational considerations.

The main contribution of this burgeoning literature is that it identifies an independent panel effect that cannot be reduced simply to the aggregation of personal preferences on a panel. The key insight is that if the racial (or gender) composition of panels had no *independent effect* on judges' rulings, then racial (gender) diversification would not lead to observable changes in judicial outcomes since minority (female) judges rarely constitute a majority in three-judge panels.

The existing literature identifies several reasons that panel composition might exert an independent effect on judicial decision making. These factors are expected, on average, to lead mixed-race, multi-judge panels to favor the interests of minority litigants and defendants more often than what would be implied by a simple majority-decision rule.

Voting Effects. Due to institutional considerations—and to avoid the need to write costly dissenting opinions—judges may change their vote simply because of how their co-panelists vote. The idea is that courts often follow norms such as consensus and deference to seniority, which limit discretion and reduce the likelihood that judges will dissent (Posner 2008). See Fischman (n.d.) for a model that accounts for how institutional norms contribute to the fact that one judge's vote is likely endogenous to those of his colleagues.

Deliberation Effects. Influenced by their colleagues' reasoning, judges may vote differently than what may be expected based on their ascriptive characteristics, ideology, or past voting record (Eisenberg, Fisher, and Rosen-Zvi 2013). This is the case when judges reevaluate their position following interactions with judges who provide new information or insight based on their different points of view or life experiences. Minority representation might affect judicial outcomes not only through deliberation among panelists, but also by affecting how the case is presented and argued by litigants, defendants, prosecutors, and defense attorneys (Anwar, Bayer, and Hjalmarsson 2012).[6]

Presence Effects. Judges from majority groups may change their vote in the presence of judges from disadvantaged groups, particularly on cases that are known to be salient to members of the disadvantaged group. This can be due to the informal influence of collegiality (associated with dissent aversion) or to an aversion to being identified as preventing minorities from realizing their interests in highly salient matters.

Consistent with the above framework, two recent studies have found that racial diversity has significantly affected the outcomes of civil rights cases in U.S. appellate courts. Cox and Miles (2008) find that the presence of an African American judge increased the likelihood that three-judge panels ruled that the Voting Rights Act had been violated. Similarly, Kastellec (2013) finds that racially mixed panels are more likely to vote in favor of an affirmative action program.[7] Importantly, however, both Cox and Miles (2008) and Kastellec (2013) readily admit that since civil rights issues are highly salient to the black community in the United States, they represent "best case scenarios" for finding race-based panel effects in favor of minorities' interests. When racial issues are clearly at stake, the above factors—voting, deliberation, and "presence" effects—all influence more favorable outcomes for members of the disadvantaged group.

One cannot automatically deduce that mixed panels work in favor of minority defendants in criminal appeals from the fact that they influence the outcomes of civil rights cases. This remains an open empirical question. As shown earlier, in single-judge settings, ingroup favoritism is least likely—but not unlikely—in criminal sentencing. However, it is possible that multijudge settings decisively alter how minority judges trade off between identity and institutional considerations in criminal sentencing. For example, the fact that the judicial outcome is attributed to the panel, rather than to individual judges, may reduce some of the disincentives for minority judges to advocate for (or simply vote in favor of) more lenient treatment of minority defendants. It is, thus, plausible that mixed panels will be more likely to rule in favor of minority

[6]Note that institutional factors—for example, the unanimity norm or logrolling—implicitly suggest that judges are strategic and do not necessarily vote according to their "sincere" preferences (Fischman n.d.). By contrast, deliberation explanations implicitly assume sincere voting.

[7]These findings corroborate three studies that find gender-based panel effects in sexual harassment and sex discrimination cases: Farhang and Wawro (2004), Peresie (2005), and Boyd, Epstein, and Martin (2010).

defendants in criminal appeals than panels comprised exclusively of members of the dominant group. In the next section, we describe the institutional setting in which we test this hypothesis.

The Political and Judicial Context

Israel offers an interesting context for examining the relationship between judicial diversity and the outcomes of the justice system. The Arab citizens of Israel constitute a distinct national minority, making up about 20% of the population. Relations between the Jewish majority and Arab minority since Israel's founding in 1948 have been marked by tension with varying degrees of intensity. Given Israel's conflictual relationship with neighboring Arab states, the country's Arab citizens were initially considered a security risk. Between 1948 and 1966, Arab Israelis lived under military administrative rule, which had far-reaching implications for their potential social and economic mobility. The abolition of military rule and the removal of formal restrictions did not, however, eliminate informal forms of ethnic inequality, partly stemming from tensions between Israel's democratic ideals and its national Jewish identity (Gavison 1999).

Though Arab Israelis (unlike Palestinians in the West Bank and Gaza Strip) formally enjoy most of the social and political rights available to Jewish citizens, the relationship between the two communities is still marred by a fair degree of hostility. Arab Israelis have relatively low socioeconomic status, are discriminated against in land allocations, and receive smaller per capita budget allocations for health, education, and municipal infrastructure (Peleg and Waxman 2011). Discrimination against Arabs in the private labor and housing markets is also pervasive (Khaṭṭāb and Miaari 2013).

Israel's Judicial System

The Israeli legal system is based on common law due to its roots in the legislation of the British Mandatory regime, but it has no juries; the lower courts are composed of professional judges. Israel's court system has three tiers: magistrate courts (located in most major towns), district courts (located in Jerusalem, Tel Aviv, Petach Tikva, Nazareth, Haifa, and Be'er Sheva), and the Supreme Court in Jerusalem. All courts have the authority to deal with both criminal and civil matters, though the jurisdiction of magistrate courts is limited to relatively minor criminal offenses for which the accused faces up to seven years of imprisonment. The district courts serve as courts of first instance for severe criminal matters and as appellate courts for magistrate court rulings.[8] Three-judge panels hear almost all appeals from magistrate courts. Finally, the Supreme Court acts as a further appellate court and hears both criminal and civil cases.

The state and the defendant may appeal the verdict, the sentence, or both. The case law specifies that appellate courts should defer to the sentence imposed by the lower court and alter it only when it is "unreasonable" and when it deviates "extremely and substantially" from proper sentencing. Appellate courts, however, alter sentences quite frequently. In our database, the appellate court altered sentences in about 53% of appeals.

Judicial Diversity in Israel's Courts

One of the defining features of Israel's court system is its appointment procedure, which was designed to increase the court's independence. According to Israel's Basic Law, a nonpartisan selection committee appoints judges, who do not have to be confirmed by the Knesset. The committee is composed of nine members: two cabinet members, one of whom is the justice minister; two Knesset members, one of whom is a member of the coalition and one whom is an opposition MP; two members of the Israel Bar Association; and three Supreme Court judges, one of whom is the chief justice. In practice, it is difficult (but not impossible) to nominate a candidate without the consent of the Supreme Court members of the committee, who tend to vote as a bloc. Importantly, the composition of the committee ensures that hiring and promotion decisions cannot be captured by one group of decision makers. Once nominated, judges enjoy a secured tenure until their mandatory retirement age of 70.

This selection process has important implications. On the one hand, judges are generally considered to be nonpartisan. In fact, it is rarely possible to infer the ideology of a judge in Israel from the identity of the political bloc that is in power during her confirmation (Weinshall-Margel 2011, 563). Partly because judges are appointed, for the most part, according to merit rather than political affiliation, the judicial system is generally held in high esteem by the Israeli public (Fishman, Rattner, and Turjeman 2006). On the other hand, the selection procedure is not conducive to diversity, a fact that has been a source of much debate in recent years.[9] Notably, notwithstanding many attempts to reform the

[8] With few exceptions, serious criminal offenses are tried by a panel of three judges at the district level, with an automatic right to appeal to a multijudge panel at the Supreme Court.

[9] For example, at the foundation of the state of Israel in 1948, all Supreme Court judges were Jewish, male, and Ashkenazi. The first

selection process to make the judiciary more representative, the professional bureaucracy at the Ministry of Justice rejects the idea of a "representative" court.

Rejecting descriptive representation on the bench, guidelines to judges' nominations stipulate that the court should ideally reflect the country's diversity, but the pool of candidates must be determined first and foremost by merit and commitment to the values of the state of Israel (Zamir, Rubinstein, and Guy-Ron 2001). Given the objection to descriptive representation, and the structural inequalities in Arabs' access to higher education, it is hardly surprising that the Arab minority in Israel is underrepresented in the judiciary. For example, the share of Arab judges in the district court system has remained fairly constant over the past three decades, hovering around 4–6%, about one-third of the share of Arabs in the population (authors' calculation).[10] To that extent, the findings of this study bear directly on the ongoing debate in Israel and elsewhere on the implications of diversifying the courts.

Disparities in the Israeli Justice System

To what extent do Israel's political and social realities permeate the courts? Like minority groups in many other countries, Arabs are overrepresented in prisons. This, however, can be the result of either systemic inequities or differences in criminal behavior. Early studies often reported that Arabs were treated more harshly than Jews by the justice system (Cohen, Fishman, and Soroka 1985; Cohen and Palmor 1985; Kretzmer 1990). However, as elsewhere, a strand of the early literature suggests that when controlling for additional factors such as criminal record, pretrial detention, and sentencing requests, the evidence of overt judicial discrimination is more ambiguous (Haj Yahia, Rahav, and Teichman 1994; Hassin and Kremnitzer 1993).

In one of the most comprehensive studies of discrimination in the criminal justice system in Israel, Rattner and Fishman (1998) find that the system is consistently harsher toward Arabs, at least in the period covered by their study (1980–92). Compared to Jews with similar demographic and case characteristics, Arabs had a higher probability of being convicted and sentenced to prison. More recently, Fishman, Rattner, and Turjeman (2006) report similar findings using information on all violent offenses between 1985 and 2000 in two district courts. These studies suggest that there is a wide discrepancy between the sentencing of Jews and Arabs.

Unlike past work on disparities in criminal court systems, our main goal is not to estimate ethnic gaps in punishment. This is because controlling for even a large number of defendant and case characteristics is usually insufficient for causal identification, since differences in case outcomes may be due to unobserved, omitted case characteristics that are *correlated with ethnicity* (Abrams, Bertrand, and Mullainathan 2012). Instead, we seek to take advantage of the randomness in the assignment of panels to cases in order to estimate the causal effects of mixed panels, *conditional on the defendant's ethnicity*. In so doing, any unobserved characteristics correlated with ethnicity are controlled for, and the causal effect is measured for an input variable that can be manipulated (panel composition) rather than one that is fixed (defendant ethnicity).

Panel Effects in Sentencing Appeals in Israel

Following Kastellec (2013), Arab judges may affect panel decision making through the three mechanisms described above: voting, deliberation, and presence. For example, Arab judges may influence their Jewish colleagues by voting, or arguing, for more lenient outcomes for Arab defendants. The third mechanism is also plausible in sentencing appeals: The mere presence of an Arab judge likely increases the salience of the defendant's ethnicity compared to cases decided by an all-Jewish panel. This exogenous increase in salience may cause members of the majority group to treat minority defendants less harshly, even if the minority judge makes no explicit plea (Anwar, Bayer, and Hjalmarsson 2012; Sommers 2006).[11] In the next section, we empirically test whether panel effects that have been identified in civil rights cases extend to criminal cases. If we find evidence of panel effects in criminal cases, the dynamics of judicial behavior in multi-judge settings may be much more significant than previously thought.

Data and Estimation Strategy

The data used in this article have been assembled by the authors using the Nevo database (www.nevo.co.il), which

non-Ashkenazi (Sephardi) Supreme Court judge was appointed in 1962, the first woman in 1977, and the first Arab in 2004.

[10]Compare this to the United States, where the share of African American judges in federal courts is now almost identical to the share of African Americans in the population (Kastellec 2013).

[11]Our theoretical argument is premised on the idea that there is inherent asymmetry between majority and minority groups. Majority ethnic groups, in Israel and elsewhere, tend to be "ethnic blind" (just as men tend to be "gender blind"). As such, ethnicity is less likely to be salient for Jewish defendants as compared to a minority Arab defendant. This means that the mechanisms that underline panel effects, such as the presence of an Arab judge, are much more likely to be relevant for Arab defendants.

Table 2 Appeals by Court and Ethnicity (All District Courts)

	Number of Arab Judges				Population	
Court	**0**	**1**	**2**	**Total**	**100k**	**Share**
Jewish Defendant						
Nazareth	6	88	16	110	573	0.10
Haifa	328	2	0	330	642	0.11
Tel Aviv	280	61	4	345	1,214	0.22
Petach Tikva	371	0	0	371	1,666	0.30
Jerusalem	152	7	0	159	650	0.12
Be'er-Sheva	137	0	0	137	844	0.15
Total	1,274	158	20	1,452	5,558	1
Arab Defendant						
Nazareth	6	117	25	148	694	0.43
Haifa	239	1	0	240	233	0.14
Tel Aviv	75	6	0	81	18	0.01
Petach Tikva	121	0	0	121	156	0.10
Jerusalem	100	8	0	108	303	0.19
Be'er-Sheva	48	0	0	48	206	0.13
Total	589	132	25	746	1,610	1
National Total	1,863	290	45	2,198	7,168	

Source: Nevo database (in Hebrew). Population figures for Israeli Jews (top panel) and Israeli Arabs (bottom panel) are derived from Israel's Central Bureau of Statistics (www.cbs.gov.il/reader).

collects decisions directly from the computerized court system.[12] We first retrieved all available final decisions in criminal appeals cases between 2007 and 2011 in Israel's six district courts and manually coded the ethnicity of the defendants and panel judges. We coded defendants as Arabs or Jews based on their names and coded judges' ethnicity based on their names and the biographical information found on the Israeli Judicial Authority official website (www.court.gov.il/heb/home.htm).[13] The distribution of the entire set of appeals by the number of Arab judges on a panel is reported in Table 2. Nationally, out of 1,863 appeals, 335 (18%) were heard by at least one Arab judge, though this share is larger for Arab defendants (26%) than for Jewish defendants (14%), reflecting the relative large number of Arab judges in Nazareth district, which covers a large share of Israel's Arab population.

In the second stage, we focus on manually coding cases from three district courts: Tel Aviv, Jerusalem, and Nazareth ($n = 951$), which have jurisdiction over an area that includes 63% of Israel's Arab population and 44% of its Jewish population. We did not collect further data from Petach Tikva, Be'er-Sheva, or Haifa district courts, since they had no Arab judge hearing criminal appeals during this period.[14] Criminal appeals solely against convictions and those that were heard by a single judge (almost exclusively traffic offenses) were then excluded, which left 594 sentence appeal decisions by a panel of three judges from three courts.

Given this relatively small sample size, we use the Amelia II package in R (Honaker, King, and Blackwell 2011) to impute missing values for the independent variables—but not the dependent variables—used in the empirical analysis. We do not, however, impute missing values for the 47 observations that had more than two

[12]The Nevo database is the most comprehensive legal database in Israel, containing about 2.8 million cases from the magistrate courts, district courts and Supreme Court as of October 2013.

[13]Though simple, our method of ethnic identification is highly reliable. For example, Shayo and Zussman (2011) analyze the relationship between names and ethnicity using Israel's Population Registry and find that 63% of first names are exclusively Jewish and 28% are exclusively Arab. When adding family name, the share of population that can be exclusively classified as Jewish and Arab rises to 95% (with the rest mostly being a subset of immigrants from the former Soviet Union who are not formally classified as Jewish).

[14]In Jerusalem, the Arab judge who served on criminal appeals retired in mid-2009; thus, data were collected only for 2007–09. Similarly, for Tel Aviv district court, data were drawn only for the 2008–11 period because no Arab judge had participated in criminal appeals panels in 2007. We exclude Haifa even though it had an Arab judge appointed temporarily for a short period in 2009, who heard three cases.

missing values for any of the eighteen independent variables we use in our regression models. Three additional observations were dropped since we did not have information on the defendant's name (and thus were unable to assign ethnicity), which reduced our sample size to 544. Importantly, the results reported in the article are robust to the exclusion of observations with any missing values.

Identification Strategy

Our identification strategy relies on random features in the assignment of judges to cases. First, court clerks assemble panels based on the judges' schedule availability and their expertise. Some judges are, thus, more likely to be assigned to criminal appeals panels, whereas others are more likely to be assigned to other judicial tasks (e.g., committees) or other types of cases (e.g., civil rights, tort, or administrative law). Importantly, there is nonetheless sufficient rotation among judges to ensure many different realizations of three-judge panels. This sort of availability-based rotation is what ensures that while the assignment process is not fully randomized, there is sufficient variation in the ethnic composition of the three-judge panel that is plausibly random.

Second, after being assembled by the court clerk, panels are assigned to hear criminal appeals on specific dates. The key point is that the clerk assigns cases to panels based on caseload considerations (i.e., the order in which appeals enter the docket), and the availability of judges is independent of the cases on the court's docket.[15] To sum up, defendants are assigned a date of appeal (and thus enter the court docket) independently of the assignment of judges to serve on a specific panel on a specific date. This ensures that both Arab and Jewish defendants can find themselves in front of a panel that has characteristics (e.g., age, gender, and ethnic composition) that are orthogonal to the characteristics of the case. So although Arab judges may be overrepresented in hearing criminal cases, this process can be considered (at least) quasi-random in that it ensures orthogonality between case characteristics and panel characteristics within courts. Tables A2 and A3 in the appendix provide evidence in support of our identification assumptions.

Description of Key Variables

The study's key independent variable is *Mixed Panel*, which takes a value of 1 if at least one Arab judge is on the panel, and 0 otherwise. The total number of judges in our data set is 87, of whom 80 are Jewish. Even though the share of Arab judges in the three included district courts is 9%, panels that included at least one Arab judge decided 56% of the cases. Note that the high percentage of mixed-panel cases is a function of our inclusion criteria—as indicated, we did not include cases that were heard by an all-Jewish panel when there were no Arab judges hearing criminal cases in that court during that year.[16] Also note that Jewish and Arab defendants are not equally likely to face a mixed panel (Table 2) because Arab judges and appeals by Arab defendants are unequally distributed across the courts. This distribution has important implications for our estimation strategy, which we discuss below.[17]

Turning to our dependent variables, past studies commonly use incarceration rates and sentence severity as the key outcome measures. Yet these outcomes are subject to some identification challenges that are rarely explicitly discussed. While some variables affecting sentence outcomes can be controlled for, an omitted variable bias may still be present across types of defendants and judges (Gazal-Ayal and Sulitzeanu-Kenan 2010). In this study, we introduce two alternative outcome measures—*leniency* and *harshness*—to supplement the analysis of incarceration rates and prison term lengths. Leniency is an indicator variable that takes a value of 1 when an appeal against the harshness of a sentence by a *defendant* has been accepted and a more lenient punishment has been handed down by the district (appellate) court compared to the magistrate court. Harshness is also an indicator variable, which takes a value of 1 when an appeal by the *prosecution* against the leniency of the sentence has been accepted and a harsher punishment has been handed down by the district court. Because both leniency and harshness measure *change* relative to the sentence imposed by the magistrate court, they essentially hold all case and defendant characteristics constant.

Indeed, by measuring leniency and harshness, we can overcome many of the identification challenges of previous studies. For example, if defendants facing all-Jewish panels have committed, on average, more severe offenses

[15] Conversations we had with court officials suggest that the process of allocating cases to panels is rather technical.

[16] Our inclusion criteria mechanically inflate the share of cases heard by mixed panels in our data set compared to the share of mixed panels in Israel. We used these criteria since our identification strategy is based on the counterfactual assumption that appeals that were allocated to an all-Jewish panel could have potentially been allocated to a mixed panel.

[17] To be clear, in Israel, criminal appeals are not equally allocated between the judges in the court. Some judges might hear more criminal appeals at a certain time, whereas others concentrate more on other tasks in the court and hence are less frequently allocated to criminal cases. Yet when a panel is formed to hear several criminal appeals on a specific day, the allocation of the cases to this panel is random and is usually done by the court secretariat based on managerial considerations.

TABLE 3 Leniency by Defendant Ethnicity and Panel Composition

	Defendant Appeals with More Lenient Verdict (%)		
	Jewish	**Arab**	**Total**
All-Jewish Panel	39.4	24.0	177
Mixed Panel	36.2	35.6	273
Total	268	182	450

Note: Defendant appeals (n = 450).

TABLE 4 Harshness by Defendant Ethnicity and Panel Composition

	Prosecutorial Appeals With Harsher Verdict (%)		
	Jewish	**Arab**	**Total**
All-Jewish Panel	78.6	82.9	91
Mixed Panel	79.3	62.5	45
Total	85	51	136

Note: Prosecution appeals (n = 136).

than those facing mixed panels, and these differences are not captured by our control variables, such unobserved differences should be reflected in the sentences imposed by the lower court. They should not, however, result in a difference in the probability of winning an appeal. Since the magistrate and district courts are required to take into account the same factors in sentencing, a district court decision to override the lower court's sentencing is unlikely to be explained by the offense severity or some other case-specific characteristics. In other words, by examining the acceptance rates of appeals, the magistrate court decision serves as a control for all of the factors that should affect sentencing, leaving the differences between the attitudes of the two courts as a likely explanation for the result.

Moreover, in studies examining the effect of judges' identities, studying appeal acceptance rates may be more relevant than incarceration rates because decisions regarding who should win an appeal cover a much larger range of cases than those pertaining to incarceration. In many cases, when the offense is severe (lax), there is no real dispute over whether the sentence must include (exclude) imprisonment. In such cases, the characteristics of judges and defendants do not affect incarceration rates, which may explain why past studies find rather weak effects of race in criminal cases. By contrast, in all appeals, judges must either accept or reject the appellant's position that the lower court was too harsh. Even if the appellate court imposes a prison sentence, the decision might still favor the defendant in the event of commutation. Thus, examining the identity of appeal winners is arguably a more precise way of identifying judicial biases.

Leniency and Harshness: Analysis of Raw Data

First, we consider cases in which the defendant has filed an appeal against the severity of the sentence. Table 3 provides information on the rate at which district courts handed down more lenient sentences than the lower courts, conditioning on the defendant's ethnicity and the panel composition. While the rate of successful appeals of Jewish defendants is similar in all-Jewish and mixed panels, Arab defendants have an almost 50% greater chance of winning an appeal when facing a panel with at least one Arab judge.

Now consider a similar analysis for the subset of cases in which the prosecution appealed the sentence, looking this time at the rate at which the appellate court assigns *harsher* sentences (Table 4). Again, Jewish defendants see similar rates of harshness in both types of panels, whereas Arab defendants enjoy a large decrease in the likelihood of receiving a harsher sentence when facing a mixed panel. In the next section, we turn to regression analysis to more rigorously examine the significance of these findings.

Estimation Strategy

Although the assignment of panels to cases within courts is quasi-random, the fact that mixed panels are disproportionately located in Nazareth (where there are more Arab defendants), whereas homogeneous panels are in the other two district courts (where there are more Jewish defendants), creates some challenges for estimation. Concretely, there are likely differences across the district courts that could be mistakenly interpreted as evidence of a treatment effect of mixed panels. The important point for our analysis is that these differences in the covariates of interest across treatment and control are balanced, *conditional on the district court in which the case is being heard.* Tables A2 and A3 in the appendix provide strong evidence that our identification assumption holds. With the exception of very few covariates, almost all variables are balanced between all-Jewish and mixed panels, conditional on defendant ethnicity and court fixed (i.e., unobserved) effects.

To estimate the impact of mixed panels on the probability of receiving a more lenient sentence, we thus fit the following logistic model:

$$Pr(y_i = 1|x_{1i}, x_{2i}, X_i, F) = logit^{-1}(\alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} * x_{2i} + \beta X_i + \gamma F_i) \quad (1)$$

where y_i is the indicator variable *leniency*; x_{1i} is an indicator of the respondent's ethnicity (which takes a value of 1 for Arab defendants and 0 for Jewish defendants); x_{2i} is the *mixed-panel* indicator that takes a value of 1 if at least one of the panel judges is Arab, and 0 for all-Jewish panels. All models include an interaction term between the defendant's ethnicity (x_{1i}) and the panel composition (x_{2i}), a vector of controls (X_i), and indicators for Tel Aviv and Jerusalem district courts (F_i), with Nazareth serving as the base category. The district court fixed effects are needed to soak up differences across districts and to take advantage of the as-if random assignment of panels *within courts*. The leniency model was fit only using observations in which the defendant filed the appeal. Similar logistic models were used for the probability of receiving a harsher sentence (*harshness*). The harshness model was fit only using observations in which the prosecutor filed the appeal.

We fit four models for both the leniency and harshness outcomes. Model 1 is a base model that includes the key independent variables (defendant's ethnicity, panel composition, and their interaction), district court indicators, defendant's sex, and an indicator of the defendant's past criminal record. Model 2 adds judge covariates, including the number of female judges, judge age, and judicial experience (years as a judge), which are averaged across the three panel judges. Model 3 adds offense controls, including the type of offense (collapsed into four categories) and an indicator of the victim's ethnicity (1 = Jewish; 0 otherwise). Model 4 adds court procedural controls, including indicator variables for whether the defendant was convicted in the magistrate court using a plea bargain or trial, and an indicator that takes a value of 1 if the prosecution explicitly requested the court to reject the appeal. In order to account for possible correlation between residuals among cases decided by the same panel of three judges, we used cluster-robust estimates of standard error based on the Huber sandwich estimator, in which each cluster was defined by a separate unique panel.

Results

We report detailed results for the four types of models in Tables 5 and 6 and in Figure 1.

Starting with *leniency*, the regression results corroborate our finding that Arab defendants are significantly more likely to have their appeal accepted by a mixed panel compared to an all-Jewish panel. The size of this effect is about 22 percentage points using Model 4 ($p = .001$), which has the best fit (Table 5). By contrast, the likelihood that Jewish defendants will have their appeal accepted is independent of the panel's ethnic composition. Moving to *harshness*, we find that Arab defendants are also much less likely to receive a harsher punishment after the prosecution appeals if the appeal is heard by a mixed panel compared to an all-Jewish panel. Although our best model shows the effect at 31.5 percentage points, robustness analysis (described below) suggests it may be closer to 20 or 25 percentage points. Though the point estimates are very large, these results fall somewhat below the 95% significance level ($p = .093$) due to the relatively small sample size ($n = 136$). Again, the likelihood that Jewish defendants will receive a harsher punishment following the prosecution's appeal is independent of the panel's ethnic composition.

Together, the above findings suggest that Arab defendants receive more favorable treatment from mixed panels than from all-Jewish panels. As mentioned, the advantage of estimating *leniency* and *harshness* is that they allow us to account for unobservable case and defendant characteristics since they estimate change in relation to the magistrate court. The disadvantage of the above analysis is its limited ability to reveal the extent to which deferential treatment by types of panels translates into actual sentencing. In the next section, we therefore analyze two additional outcome variables: *incarceration* and *prison term*.

Incarceration

Incarceration is a binary variable that equals 0 if the defendant was not sentenced to prison and 1 for all positive prison sentences, using the full data set ($n = 544$). The raw data (Table 7) suggest that the appellate court's decision to incarcerate Jewish defendants is independent of panel composition. By contrast, the likelihood that Arab defendants will serve a prison term is 19 percentage points lower when their case is heard by a mixed panel compared to an all-Jewish panel.

To examine the significance of this finding, we run a series of logistic regressions as described above in Equation (1). Corroborating the results of Table 7, we find that the marginal effect of mixed panels is negative for Jewish defendants but insignificant ($p = .867$) and significant for Arab defendants ($p = .005$). As Table A4 in the appendix and Figure 2 show, across all four models, we find that Arab defendants have a 17–18% lower probability of being incarcerated when their appeal is heard by a mixed panel. The robustness analysis in the next main section gives similar estimates in the range of 14–20%. By contrast, incarceration rates for Jewish defendants are independent of the panel's ethnic composition. The robustness of these findings across all models provides

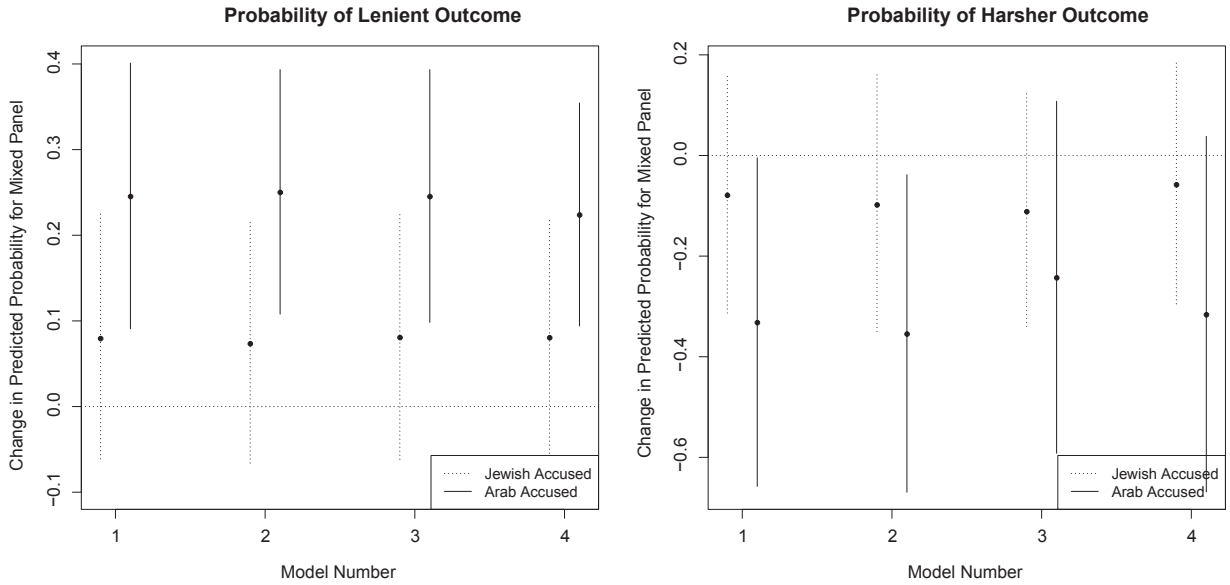
TABLE 5 Lenient Higher-court Verdict Rates for Defendant Appeals (Full Logistic Model)

	Model 1			Model 2			Model 3			Model 4		
More Lenient	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Arab Defendant	−0.781	(0.433)	.072	−0.806	(0.42)	.055	−0.765	(0.43)	.076	−0.723	(0.444)	.104
Mixed Panel	0.358	(0.284)	.208	0.35	(0.322)	.279	0.388	(0.354)	.275	0.456	(0.441)	.302
Mixed Panel × Arab Defendant	0.983	(0.454)	.031	1.103	(0.453)	.015	1.039	(0.452)	.022	1.042	(0.477)	.029
Nazareth Court	−0.693	(0.352)	.05	−2.836	(0.567)	0	−2.868	(0.6)	0	−2.21	(0.676)	.001
Jerusalem Court	0.151	(0.294)	.607	−1.224	(0.506)	.016	−1.262	(0.533)	.018	−1.235	(0.59)	.037
Female Defendant	1.06	(0.433)	.015	1.187	(0.484)	.015	1.204	(0.497)	.016	1.426	(0.499)	.005
Previous Criminal Record	0.324	(0.261)	.215	0.275	(0.262)	.294	0.383	(0.276)	.165	0.343	(0.28)	.22
Average Judge Age				0.008	(0.063)	.899	0.009	(0.062)	.884	0.063	(0.063)	.316
Average Judge Experience				−0.209	(0.075)	.005	−0.216	(0.074)	.004	−0.253	(0.083)	.003
Female Judges				0.15	(0.238)	.528	0.142	(0.239)	.551	0.175	(0.247)	.479
Physical or Sexual Assault							−0.389	(0.225)	.085	−0.569	(0.235)	.016
Property or Fraud							−0.214	(0.19)	.261	−0.35	(0.2)	.082
Fiscal, Economic, or Business							−0.048	(0.373)	.899	0.097	(0.353)	.783
Regulatory							−0.2	(0.464)	.667	−0.088	(0.564)	.877
Jewish Victim							−0.096	(0.241)	.691	0.063	(0.217)	.773
Prosecution Requested Rejection of Appeal										−1.584	(0.507)	.002
Convicted by Guilty Plea										−0.295	(0.23)	.2
Convicted by Trial										1.211	(0.538)	.025
(Intercept)	−0.759	(0.24)	.002	3.86	(2.925)	.188	4.214	(2.953)	.154	2.373	(2.703)	.381
Number of Predictors	7			10			15			18		
AIC Value	584.537			568.681			575.597			518.302		
Observations	450			450			450			450		
Marg. Effect of Mixed Panel (Jewish Def.)	0.08	(0.074)	.279	0.074	(0.072)	.304	0.081	(0.073)	.266	0.081	(0.069)	.244
Marg. Effect of Mixed Panel (Arab Def.)	0.246	(0.079)	.002	0.251	(0.073)	.001	0.246	(0.075)	.001	0.224	(0.066)	.001

TABLE 6 Harsher Higher-court Verdict Rates for Prosecutorial Appeals (Full Logistic Model)

	Model 1			Model 2			Model 3			Model 4		
Harsher	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**	**Coefficient**	**SE**	**p-value**
Arab Defendant	0.461	(0.731)	.529	0.409	(0.768)	.595	0.532	(0.893)	.552	0.406	(0.951)	.67
Mixed Panel	−0.475	(0.728)	.516	−0.662	(1.032)	.522	−0.917	(1.002)	.362	−0.52	(1.044)	.619
Mixed Panel × Arab Defendant	−1.307	(0.91)	.154	−1.44	(1)	.153	−0.745	(1.119)	.507	−1.827	(1.506)	.228
Nazareth Court	0.099	(0.712)	.89	0.556	(1.311)	.672	−0.339	(1.438)	.814	−0.404	(1.896)	.831
Jerusalem Court	−0.909	(0.723)	.211	−2.083	(1.198)	.085	−2.629	(1.254)	.038	−3.29	(1.441)	.024
Female Defendant	1.67	(5.222)	.75	1.908	(4.217)	.652	1.674	(5.217)	.749	1.567	(4.748)	.742
Previous Criminal Record	0.44	(0.68)	.519	0.745	(0.756)	.326	0.773	(0.839)	.359	0.892	(0.862)	.303
Average Judge Age				0.006	(0.158)	.972	−0.042	(0.143)	.771	−0.09	(0.169)	.597
Average Judge Experience				0.1	(0.187)	.593	0.194	(0.183)	.294	0.126	(0.208)	.548
Female Judges				−1.361	(0.575)	.02	−1.772	(0.692)	.012	−1.841	(0.743)	.015
Physical or Sexual Assault							0.187	(0.886)	.833	−0.055	(1.211)	.964
Property or Fraud							−0.206	(0.908)	.82	−0.061	(1.066)	.954
Fiscal, Economic, or Business							0.447	(1.01)	.659	0.621	(1.354)	.648
Regulatory							1.475	(0.938)	.118	1.656	(1.302)	.206
Jewish Victim							3.243	(1.492)	.032	3.725	(1.841)	.045
Prosecution Requested Rejection of Appeal										−4.336	(1.361)	.002
Convicted by Guilty Plea										−0.08	(0.889)	.928
Convicted by Trial										−1.208	(1.177)	.307
(Intercept)	1.64	(0.658)	.014	1.341	(7.81)	.864	2.11	(7.226)	.771	7.461	(9.029)	.41
Number of Predictors	7			10			15			18		
AIC Value	150.68			146.144			136.731			133.081		
Observations	136			136			136			136		
Marg. Effect of Mixed Panel (Jewish Def.)	−0.078	(0.12)	.517	−0.097	(0.131)	.461	−0.111	(0.119)	.355	−0.057	(0.123)	.643
Marg. Effect of Mixed Panel (Arab Def.)	−0.331	(0.167)	.049	−0.354	(0.161)	.03	−0.242	(0.178)	.178	−0.315	(0.18)	.083

FIGURE 1 Panel Effects: Harshness and Leniency



Note: The figure shows changes in the predicted probabilities of a more lenient sentence (where the defendant appealed) or a harsher sentence (where the prosecutor appealed) for defendants facing a mixed panel in each of the fitted models. The changes in predicted probability are calculated by averaging the changes in probability estimated by the model for each defendant (of the given ethnicity) in the data set at the observed values of the covariates for this defendant. Lines show 95% confidence intervals. Raw predicted probabilities from the fourth model for leniency and harshness are shown in Tables A6 and A7 in the appendix.

a strong indication of conditional ethnicity-based panel effects on incarceration rates in Israel.

The incarceration model also allows us to tentatively investigate the criminality gap between Arabs and Jews. Controlling for defendant, case, and procedural variables, all-Jewish panels are about 13% more likely to incarcerate an Arab defendant than a Jewish defendant with similar characteristics. This gap between Arab and Jewish defendants all but vanishes in the case of mixed panels. As mentioned, these findings are less robust since there are possible unobserved covariates correlated with the defendant's ethnicity.

Prison Term

To measure the length of prison terms, we subset the data to only include positive sentence terms, as zero-sentence terms are captured by the incarceration variable ($n = 379$). To estimate the impact of mixed panels on the prison term, we fit the following ordinary least squares (OLS) model:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} * x_{2i} + \beta X_i + \gamma F_i \quad (2)$$

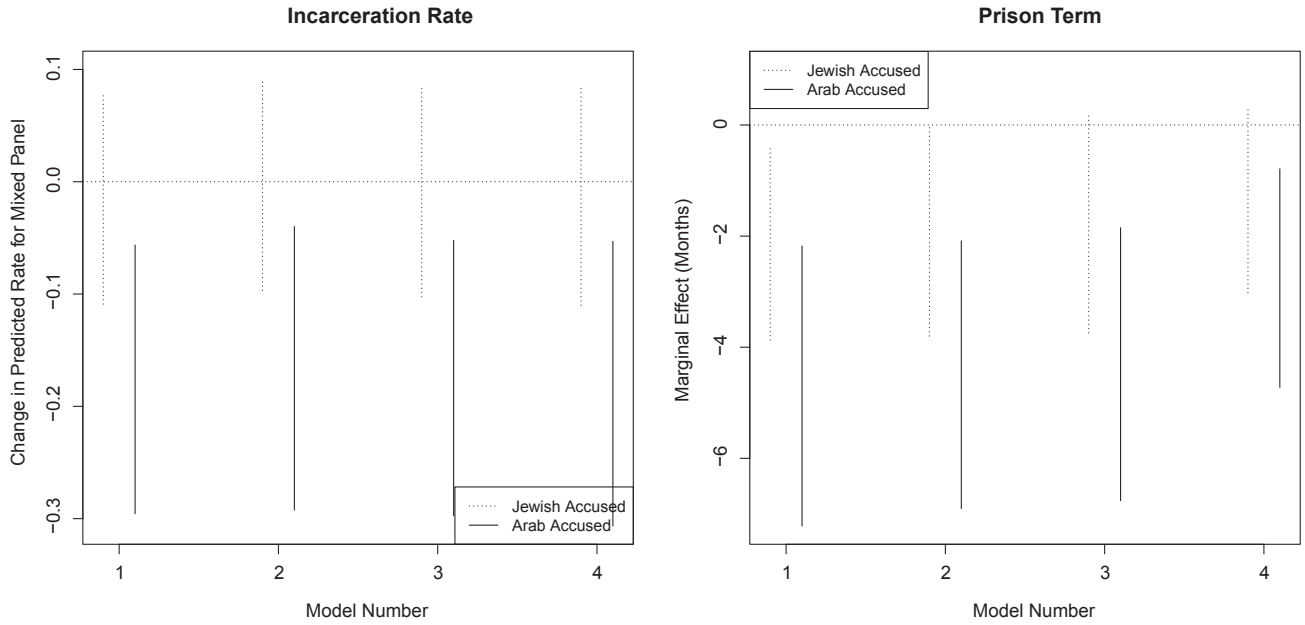
Here, y_i ($y_i > 0$ for all i) represents the prison term in months. The prison term models use the same covariates as the leniency model, described above, and include the prison term assigned at the magistrate level as an additional control variable. In all regression models, standard errors are clustered at the panel level.

Results are presented in Table A5 in the appendix and Figure 2. For Jewish defendants, the marginal effect of a mixed panel is slightly negative, but it does not reach conventional levels of statistical significance for all models. By contrast, for Arab defendants, the marginal effect of mixed panels is negative and significant, representing a decrease of 2.7–4.7 months in prison compared to all-Jewish panels ($p = .006$). The robustness checks in the next section occasionally give smaller point estimates, but they always show a negative and significant mixed-panel effect. Since the mean prison term for incarcerated Arab defendants is 18 months, the marginal effect of mixed panels is a 15–26% decrease in prison time. This finding is especially striking given the lower incarceration rate of Arab defendants by mixed panels, indicating that mixed panels hand down prison term sentences on more severe cases, on average.

Robustness

To ensure the robustness of our findings, we fit several alternative models for our outcomes of interest. Each

FIGURE 2 Panel Effects: Incarceration and Prison Term



Note: The figure shows changes in predicted probabilities of incarceration due to mixed panels and the marginal effects of mixed panels on the length of prison terms for each of the fitted models. The prison term model is an ordinary linear regression, so the effects for that model are constant. For the incarceration model, changes in predicted probability are calculated by averaging the changes in probability estimated by the model for each defendant (of the given ethnicity) in the data set at the observed values of the covariates for this defendant. Lines show 95% confidence intervals. Raw predicted probabilities from the fourth model for the incarceration model are shown in Table A8 in the appendix.

of these tests relaxes an assumption made in the above analysis or otherwise alters the modeling strategy. First, we disaggregate data by district court and examine contingency tables similar to Tables 3–7 for the individual courts. Next, we fit parsimonious versions of each of the models of the four outcomes to ensure that our models do not include spurious covariates. A third robustness check considered whether the partial effects of covariates, in addition to defendant ethnicity, might vary with panel composition. To examine robustness to this possibility, we refit the parsimonious versions of the models but expanded each pool of covariates for possible selection to include interactions with the mixed-panel indicator. Finally, we performed a nonparametric analysis using propensity score weighting (Hirano and Imbens 2001). Using the twang package in R (Ridgeway et al. 2013), we fit a propensity score model for the probability of facing a mixed panel, weighted observations by their fitted scores, and ran logistic regressions of the outcomes of interest against the presence/absence values for mixed panel. These checks and their results, which strengthen our confidence in the robustness of our findings, are described in greater detail in the online supporting information.

Conclusion

This article examines whether the ethnic composition of appellate courts has a causal impact on the outcome of sentencing appeals in Israel. Taking advantage of exogeneity in the assignment of cases to panels, we find that the ethnic composition of district court panels is highly consequential for Arab defendants, who are more likely to win their appeals (and less likely to lose the prosecution's appeals) when a panel that includes at least one Arab judge hears their case.

The magnitude of these panel effects is sizable: a 14–20% reduction in incarceration rates and a 15–26% reduction in prison term sentencing. By contrast, the outcome of the appeals process for Jewish defendants is independent of the ethnic composition of the appellate panel. This suggests that Arab judges are not more (unconditionally) liberal, eliminating a possible alternative explanation that places less emphasis on considerations of ethnic group membership. The findings are an important contribution to the literature on judicial diversity because they demonstrate that ethnicity-based panel effects are not limited to civil rights cases, which have an explicit racial dimension. Instead, panel effects are seen even in the

much more common setting of criminal sentencing appeals. These findings suggest that panel effects may be of much broader relevance to legal decision making than was previously believed.

This article also makes a theoretical contribution to the broader literature on how judges' ascriptive characteristics impact their decision making by outlining a framework that can help explain the mixed findings of ingroup bias across a wide range of judicial settings. Building on the evidence to date, our intuition is that the different characteristics of legal procedures condition how judges trade off identity considerations, which can lead to ingroup favoritism and institutional norms and pressures that are designed to reduce idiosyncratic discretion. In particular, a number of parameters of judicial processes appear to be important, including the speed with which decisions are taken, the stakes at play, the salience of the issue for disadvantaged groups, and the amount of discretion given to judges. Moreover, we suggest that some institutional factors that may constrain judges from taking their group identity into account in single-judge criminal law settings may be less binding in multi-judge settings, in which the racial and ethnic identities of individual judges can alter the panel's strategic dynamics.

While it is unclear how generalizable these results are outside the Israeli context, our general focus on how the characteristics of judicial processes—and how they influence the way identity considerations affect decision making—can be explored in other settings. There may also be other relevant contextual features that condition the influence of identity considerations, including the type of legal family, the use of juries as opposed to judges, the level of politicization of the judicial branch, and variations in the level of integration of the racial or ethnic minority group.

Another natural next step is to further investigate the mechanisms that drive panel effects. In the Israeli context, the presence of an Arab judge might affect his or her Jewish copanelists through voting, deliberation, and/or presence effects. Teasing out the relative strength of each of these plausible channels is not possible, since contextual variables (e.g., identity influences) can hardly be separated from endogenous effects (Manski 2000). Adjudicating between voting, deliberation, and presence effects is further complicated by the fact that we are not privy to the deliberation and discussions taking place between judges, and because district courts in Israel have a very strong consensus norm.

In the online supporting information, we describe a preliminary effort to push in this direction by systematically analyzing the text of the judicial opinions that make up our data set. Though we cannot observe the deliberation or arguments among district court judges, the deliberative process may be reflected in their written opinions. The idea is that if minority judges provide a different perspective based on their life experiences and their better understanding of the consequences leading minorities to commit crimes—as assumed by deliberation arguments—then this perspective may be reflected in the arguments provided by judges to justify the outcome of the judicial process. While our text analysis provides no support for a deliberation effect, it does suggest that through the creative application of new techniques, we may be able to make progress in isolating the mechanisms driving judicial biases.[18]

TABLE 7 Incarceration by Defendant Ethnicity and Panel Composition

	Jewish	Arab	Total
All-Jewish Panel	0.72	0.79	239
Mixed Panel	0.71	0.60	305
Total	324	220	544

Finally, this study discussed the value of representativeness in institutional arrangements, in both the political and judicial arenas, in order to increase the substantive representation of minorities in multiethnic societies. This article provides robust evidence that, at least in Israel, judicial diversity results in substantively different outcomes for the minority group. This is true even in a system that prides itself on its strong institutional norms, meritocratic hiring and promotion, and organizational structures designed to ensure the fair and equal application of the law. This evidence of identity's impact on judicial decision making represents a challenge to defenders of the status quo because in the absence of more equitable representation in the Israeli judiciary, it is clear that Arab defendants are at a disadvantage for no reason other than their identity. Pressure against descriptive representation in the judiciary is likely to remain strong in Israel and beyond—given the strong commitment of many jurists to judicial independence and meritocratic promotion—but sustaining the legitimacy of judiciaries in the face of mounting evidence of racial and ethnic bias will depend on finding concrete ways to address the real concerns of disadvantaged groups. One place to start is greater transparency about how judicial identities are shaping rulings. If judiciaries are committed to

[18]Not finding evidence of a deliberation effect is not the same as finding evidence of no effect; that is, it may be that Arab judges provide a different perspective, which is not reflected in the final written opinion.

promoting the fair and equal application of the law, they must take it upon themselves to ensure that evidence of identity effects is systematically collected and publicized (perhaps by an independent group) so as to facilitate a serious conversation about the norms and practices needed to facilitate fair and equal treatment under the law for all.

Appendix

In this appendix, we begin by clearly defining the key independent and control variables before providing the full set of results in tabular form.

- *Arab Defendant*: an indicator variable that takes a value of 1 if the defendant is Arab and 0 if he or she is Jewish.
- *Mixed Panel*: an indicator variable that takes a value of 1 if the panel of three judges includes at least one Arab judge, and 0 for an all-Jewish panel.
- *Prison Term, Magistrate Court*: a continuous variable capturing the prison term length (in months) handed down by the magistrate court.
- *Tel Aviv Court*: an indicator variable that takes a value of 1 for Tel Aviv district court appeals, and 0 otherwise.
- *Nazareth Court*: an indicator variable that takes a value of 1 for Nazareth district court appeals, and 0 otherwise.
- *Jerusalem Court*: an indicator variable that takes a value of 1 for Jerusalem district court appeals, and 0 otherwise.
- *Female Defendant*: an indicator variable that takes a value of 1 if the defendant is female, and 0 otherwise.
- *Previous Criminal Record*: an indicator variable that takes a value of 1 if the defendant has a criminal record, and 0 otherwise.
- *Average Judge Age*: a continuous variable measuring the mean age of the three judges on the panel.
- *Average Judge Experience*: a continuous variable measuring the mean years of experience as a judge of the three judges on the panel.
- *Female Judges*: a variable ranging from 0 to 3 that measures the number of female judges on the panel.
- *Physical or Sexual Assault*: an indicator variable that takes a value of 1 if the defendant was convicted of a morality offense (usually sex offenses), public order offense (usually resisting arrest), or bodily harm, and 0 otherwise.
- *Property or Fraud*: an indicator variable that takes a value of 1 if the defendant was convicted of a fraud or property offense, and 0 otherwise.
- *Fiscal, Economic, or Business*: an indicator variable that takes a value of 1 if the defendant was convicted of a fiscal, economic, or business offense, and 0 otherwise.
- *Regulatory*: an indicator variable that takes a value of 1 if the defendant was convicted of a regulatory offense (usually aiding illegal entry to Israel), and 0 otherwise.
- *Jewish Victim*: an indicator variable that takes a value of 1 if the victim is Jewish, and 0 otherwise.
- *Prosecution Requested Rejection*: an indicator variable that takes a value of 1 if the prosecution explicitly asked the court to reject the defendant's appeal, and 0 otherwise.
- *Convicted by Guilty Plea*: an indicator variable that takes a value of 1 if the defendant was convicted in the magistrate court by pleading guilty, and 0 otherwise.
- *Convicted by Trial*: an indicator variable that takes a value of 1 if the defendant was convicted in the magistrate court by trial (as opposed to a plea bargain or guilty plea), and 0 otherwise.

For each of the two key dependent variables—incarceration rate and prison term length—we provide two sets of results:

1. **Parsimonious Model:** This model uses a step function, as described in the text, which includes a control variable only if its inclusion improves the model fit as measured by the AIC.
2. **Full Model:** This model includes control variables that we have reason to believe affect the variance of the dependent variable, irrespective of whether they improve the model fit.

TABLE A1 Descriptive Statistics: District Appeals Sample, 2007-2011

Variable	$\bar{x}$	s	Min	$\tilde{x}$	Max	n
Panel Characteristics						
Mixed Panel	0.6	0.5	0.0	1.0	1.0	544
Female Judges on Panel	1.3	0.7	0.0	1.0	3.0	544
Average Judge Age (years)	60.0	5.7	49.7	58.7	71.0	544
Average Judge Experience (years)	19.4	5.3	10.0	17.3	28.3	544
Defendant Characteristics						
Arab Defendant	0.4	0.5	0.0	0.0	1.0	544
Female Defendant	0.0	0.2	0.0	0.0	1.0	534
Previous Criminal Record	0.6	0.5	0.0	1.0	1.0	490
Case Characteristics						
Prison Term (months): Magistrate Court	13.0	15.4	0.0	8.0	104.0	544
Community Service (months): Magistrate Court	0.8	1.9	0.0	0.0	6.0	544
Fine (shekels): Magistrate Court	15220.4	46622.9	0.0	500.0	500000.0	541
Victim Compensation (shekels): Magistrate Court	7249.4	43764.2	0.0	0.0	750000.0	537
Sent to Prison: District Court	0.7	0.5	0.0	1.0	1.0	544
Prison Term (months): District Court	12.7	15.1	0.0	9.0	104.0	544
Community Service (months): District Court	1.0	2.0	0.0	0.0	6.0	544
Fine (shekels): District Court	18740.9	65717.2	0.0	500.0	800000.0	541
Victim Compensation (shekels): District Court	8823.6	53110.5	0.0	0.0	750000.0	539
Arab Victim	0.4	0.5	0.0	0.0	1.0	492
Felony Type						
Physical or Sexual Assault	0.5	0.5	0.0	0.0	1.0	544
Property or Fraud	0.4	0.5	0.0	0.0	1.0	544
Fiscal, Economic, or Business	0.2	0.4	0.0	0.0	1.0	544
Regulatory	0.1	0.3	0.0	0.0	1.0	544
Procedural Facts						
Convicted by Guilty Plea	0.5	0.5	0.0	0.0	1.0	534
Convicted by Trial	0.2	0.4	0.0	0.0	1.0	534
Prosecution Requested Rejection	0.6	0.5	0.0	1.0	1.0	533
Defendant Appealed	0.8	0.4	0.0	1.0	1.0	544
Prosecution Appealed	0.2	0.4	0.0	0.0	1.0	544
Courts						
Nazareth Court	0.4	0.5	0.0	0.0	1.0	544
Tel Aviv Court	0.3	0.5	0.0	0.0	1.0	544
Jerusalem Court	0.3	0.4	0.0	0.0	1.0	544

TABLE A2 Identification Assumption: Balance Test (Arab Defendants)

Arab Defendants (n = 220)	Mixed Panel (Predicted Value)	All-Jewish Panel (Predicted Value)	Difference (Δ)	p-value (OLS)
Defendant Characteristics				
Female Defendant	0.04	0.05	–0.01	.81
Previous Criminal Record	0.78	0.80	–0.02	.78
Judge Covariates				
Average Judge Age	67.25	67.13	0.12	.87
Average Judge Experience	26.57	26.61	–0.04	.94
Number of Female Judges	1.88	2.04	–0.16	.21
Offense and Victim Identity				
Physical or Sexual Assault	0.56	0.51	0.04	.73
Property or Fraud	0.17	0.38	–0.21	.09
Fiscal, Economic, or Business	0.39	0.25	0.14	.15
Regulatory	0.26	0.23	0.03	.77
Jewish Victim	0.14	0.35	–0.21	.08
Court Procedural Controls				
Prison Term, Magistrate Court	18.24	19.70	–1.46	.70
Convicted by Guilty Plea	0.48	0.60	–0.13	.29
Convicted by Trial	0.24	0.18	0.06	.44
Prosecution Requested Rejection of Appeal	0.63	0.49	0.14	.19
Defendant Appealed	0.68	0.60	0.07	.40
Prosecution Appealed	0.32	0.46	–0.14	.14

Note: Table reports results analysis that examines the plausibility of the assumption underlying this study's causal identification. Specifically, we examine the relationships between the study's key independent variable (*mixed panel*) and all available covariates, conditional on court fixed effects among cases with Arab defendants. We regressed each available covariate on the mixed-panel indicator and fixed court effects (using OLS). The first column shows a prediction from the model for an Arab defendant in Tel Aviv facing a mixed panel, the second column shows the predicted value for a similar defendant facing an all-Jewish panel, and the third and fourth columns show the difference of these predictions and its corresponding p-value (using the t-test). Nonsignificant differences indicate that there is no strong relationship between the covariate and the independent variable when conditioning on courts, as in all our regression models. Note that the third and fourth columns are independent of fixed court effects and remain the same for Nazareth and Jerusalem.

TABLE A3 Identification Assumption: Balance Test (Jewish Defendants)

Jewish Defendants (n = 324)	Mixed Panel (Predicted Value)	All-Jewish Panel (Predicted Value)	Difference (Δ)	p value (OLS)
Defendant Characteristics				
Female Defendant	0.06	0.09	–0.03	.26
Previous Criminal Record	0.50	0.56	–0.06	.41
Judge Covariates				
Average Judge Age	65.10	66.29	–1.19	.01
Average Judge Experience	25.70	25.86	–0.16	.66
Number of Female Judges	1.75	1.89	–0.14	.08
Offense and Victim Identity				
Physical or Sexual Assault	0.44	0.44	0.00	.99
Property or Fraud	0.29	0.35	–0.06	.38
Fiscal, Economic, or Business	0.35	0.31	0.04	.49
Regulatory	0.07	0.01	0.05	.07
Jewish Victim	0.46	0.44	0.02	.69
Court Procedural Controls				
Prison Term, Magistrate Court	14.92	13.21	1.71	.45
Convicted by Guilty Plea	0.29	0.47	–0.18	.01
Convicted by Trial	0.46	0.32	0.14	.02
Prosecution Requested Rejection of Appeal	0.65	0.56	0.08	.25
Defendant Appealed	0.82	0.79	0.03	.57
Prosecution Appealed	0.24	0.31	–0.06	.31

Note: Table reports results of analysis that examines the plausibility of the assumption underlying this study's causal identification. Specifically, we examine the relationships between the study's key independent variable (*mixed panel*) and all available covariates, conditional on court fixed effects among cases with Jewish defendants. We regressed each available covariate on the mixed-panel indicator and fixed court effects (using OLS). The first column shows a prediction from the model for a Jewish defendant in Tel Aviv facing a mixed panel, the second column shows the predicted value for a similar defendant facing an all-Jewish panel, and the third and fourth columns show the difference of these predictions and its corresponding p-value (using the t-test). Nonsignificant differences indicate that there is no strong relationship between the covariate and the independent variable when conditioning on courts, as in all our regression models. Note that the third and fourth columns are independent of fixed court effects and would remain the same for Nazareth or Jerusalem.

TABLE A4 Incarceration Rates (Full Logistic Model)

	Model 1			Model 2			Model 3			Model 4		
Incarceration	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
Arab Defendant	1.174	(0.327)	0	1.138	(0.319)	0	1.457	(0.381)	0	1.446	(0.388)	0
Mixed Panel	–0.147	(0.329)	.655	–0.048	(0.357)	.893	–0.087	(0.387)	.822	–0.125	(0.409)	.759
Mixed Panel × Arab Defendant	–1.195	(0.422)	.005	–1.209	(0.436)	.006	–1.325	(0.473)	.005	–1.346	(0.47)	.004
Prison Term, Magistrate Court	0.259	(0.049)	0	0.267	(0.053)	0	0.263	(0.046)	0	0.255	(0.048)	0
Nazareth Court	–0.852	(0.357)	.017	–0.352	(0.716)	.623	–0.216	(0.653)	.741	–0.591	(0.664)	.374
Jerusalem Court	–0.726	(0.318)	.023	–0.907	(0.525)	.085	–0.759	(0.504)	.132	–0.871	(0.509)	.088
Female Defendant	–1.178	(0.482)	.015	–1.166	(0.507)	.022	–0.906	(0.517)	.08	–1.001	(0.505)	.048
Previous Criminal Record	0.294	(0.284)	.302	0.297	(0.284)	.296	0.245	(0.293)	.403	0.264	(0.297)	.374
Average Judge Age				0.13	(0.055)	.018	0.139	(0.055)	.013	0.127	(0.053)	.017
Average Judge Experience				–0.072	(0.071)	.314	–0.068	(0.067)	.314	–0.074	(0.068)	.279
Female Judges				–0.176	(0.289)	.542	–0.194	(0.271)	.474	–0.167	(0.265)	.53
Physical or Sexual Assault							0.959	(0.419)	.023	0.989	(0.448)	.028
Property or Fraud							0.079	(0.361)	.828	0.09	(0.355)	.801
Fiscal, Economic, or Business							0.917	(0.395)	.021	0.919	(0.411)	.026
Regulatory							0.376	(0.6)	.531	0.393	(0.629)	.533
Jewish Victim							0.527	(0.315)	.095	0.495	(0.33)	.134
Prosecution Requested Rejection of Appeal										0.257	(0.347)	.46
Convicted by Guilty Plea										–0.099	(0.312)	.752
Convicted by Trial										–0.532	(0.454)	.242
(Intercept)	–0.374	(0.332)	.26	–6.812	(2.993)	.023	–8.493	(3.093)	.006	–7.43	(2.841)	.009
Number of Predictors	8			11			16			19		
AIC Value	418.426			418.188			417.901			421.047		
Observations	544			544			544			544		
Marg. Effect of Mixed Panel (Jewish Def.)	–0.017	(0.048)	.718	–0.005	(0.048)	.909	–0.01	(0.047)	.837	–0.014	(0.049)	.78
Marg. Effect of Mixed Panel (Arab Def.)	–0.176	(0.061)	.004	–0.166	(0.064)	.01	–0.175	(0.063)	.005	–0.18	(0.065)	.006

TABLE A5 Positive District level Prison Terms (Full OLS)

	Model 1			Model 2			Model 3			Model 4		
Prison Term (Months)	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**	**Estimate**	**SE**	**p-value**
Arab Defendant	1.819	(1.132)	.109	1.674	(1.035)	.107	1.572	(1.067)	.141	1.017	(0.941)	.281
Mixed Panel	–2.169	(0.889)	.015	–1.952	(0.971)	.045	–1.8	(1.001)	.073	–1.401	(0.854)	.102
Mixed Panel × Arab Defendant	–2.531	(1.195)	.035	–2.544	(1.143)	.027	–2.506	(1.143)	.029	–1.354	(0.939)	.15
Prison Term, Magistrate Court	0.876	(0.027)	0	0.882	(0.028)	0	0.858	(0.028)	0	0.907	(0.031)	0
Nazareth Court	0.587	(0.855)	.493	4.441	(2.095)	.035	3.851	(1.987)	.053	2.768	(1.369)	.044
Jerusalem Court	–3.047	(1.065)	.004	–1.29	(1.925)	.503	–1.363	(1.879)	.469	–1.669	(1.533)	.277
Female Defendant	–4.466	(1.594)	.005	–4.269	(1.524)	.005	–5.15	(1.532)	.001	–3.698	(1.777)	.038
Previous Criminal Record	–0.763	(0.671)	.256	–0.677	(0.614)	.271	–0.973	(0.638)	.128	–0.829	(0.581)	.154
Average Judge Age				0.211	(0.155)	.173	0.18	(0.153)	.241	0.179	(0.148)	.229
Average Judge Experience				0.148	(0.174)	.395	0.191	(0.187)	.308	0.077	(0.181)	.671
Female Judges				–0.273	(0.808)	.736	–0.32	(0.763)	.675	–0.182	(0.644)	.778
Physical or Sexual Assault							0.483	(1.046)	.645	0.462	(0.944)	.625
Property or Fraud							2.531	(0.882)	.004	1.814	(0.763)	.018
Fiscal, Economic, or Business							–0.392	(1.235)	.751	–0.52	(1.055)	.623
Regulatory							–0.046	(1.059)	.965	–0.492	(0.938)	.6
Jewish Victim							–0.439	(0.782)	.575	0.505	(0.566)	.373
Prosecution Requested Rejection of Appeal										–5.83	(1.095)	0
Convicted by Guilty Plea										0.008	(0.799)	.992
Convicted by Trial										–2.143	(1.104)	.053
(Intercept)	4.918	(0.857)	0	–12.641	(9.659)	.191	–11.737	(9.043)	.195	–5.931	(7.547)	.432
Number of Predictors	9			12			17			20		
AIC Value	1364.693			1361.486			1358.326			1283.461		
Observations	379			379			379			379		
Marg. Effect of Mixed Panel (Jewish Def.)	–2.169	(0.889)	.015	–1.952	(0.971)	.045	–1.8	(1.001)	.073	–1.401	(0.854)	.102
Marg. Effect of Mixed Panel (Arab Def.)	–4.699	(1.285)	0	–4.496	(1.23)	0	–4.306	(1.253)	.001	–2.755	(1.004)	.006

Table A6 Predicted Probabilities of Leniency for Defendant Appeal by Panel Type and Defendant Ethnicity

	Jew	Arab
All-Jewish Panel	0.34	0.17
Mixed Panel	0.42	0.39

Note: Probabilities computed conditional on defendant ethnicity, using Model 4 from Table 5.

Table A7 Predicted Probabilities of Harshness for Prosecutorial Appeal by Panel Type and Defendant Ethnicity

	Jew	Arab
All-Jewish Panel	0.80	0.85
Mixed Panel	0.76	0.51

Note: Probabilities computed conditional on defendant ethnicity, using Model 4 from Table 6.

Table A8 Predicted Probabilities of Incarceration by Panel Type and Defendant Ethnicity

	Jew	Arab
All-Jewish Panel	0.72	0.78
Mixed Panel	0.71	0.60

Note: Probabilities computed conditional on defendant ethnicity, using Model 4 from Table A4.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan. 2012. "Do Judges Vary in Their Treatment of Race?" *Journal of Legal Studies* 41(2): 347–83.

Allen, David W., and Diane E. Wall. 1993. "Role Orientations and Women State Supreme Court Justices." *Judicature* 77(3): 156–65.

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

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

Ayres, Ian, and Joel Waldfogel. 1994. "A Market Test for Race Discrimination in Bail Setting." *Stanford Law Review* 46(5): 987–1047.

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

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72(5): 1409–43.

Chew, Pat K., and Robert E. Kelley. 2006. "Unwrapping Racial Harassment Law." *Berkeley Journal of Employment and Labor Law* 27(1): 49–110.

Cohen, Ben-Zion, Gideon Fishman, and Joseph Soroka. 1985. "Judicial Discretion and Sentencing Disparity in Adult Felony Courts in Israel." *Journal of Criminal Justice* 13(2): 99–115.

Cohen, Ben-Zion, and Shulamit Palmor. 1985. "The Impact of Ethnicity on Probation Recommendations and Dispositions in Israel." *Justice Quarterly* 2(2): 197–212.

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

Eisenberg, Theodore, Talia Fisher, and Issi Rosen-Zvi. 2013. "Group Decision Making on Appellate Panels: Presiding Justice and Opinion Justice Influence in the Israel Supreme Court." *Psychology, Public Policy, and Law* 19(3): 282–96.

Epstein, Lee, and Jack Knight. 2013. "Reconsidering Judicial Preferences." *Annual Review of Political Science* 16(1): 11–31.

Farhang, Sean, and Gregory Wawro. 2004. "Institutional Dynamics on the U.S. Court of Appeals: Minority Representation under Panel decision making." *Journal of Law, Economics, & Organization* 20(2): 299—330.

Fischman, Joshua B. n.d. "Interpreting Circuit Court Voting Patterns: A Social Interactions Framework." *Journal of Law, Economics, and Organization.* Forthcoming.

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.

Gavison, Ruth. 1999. *Can Israel Be Both Jewish and Democratic: Tensions and Prospects.* Jerusalem, Israel: Van Leer Jerusalem Institute and Hakibbutz Hameuchad.

Gazal-Ayal, Oren, and Raanan Sulitzeanu-Kenan. 2010. "Let My People Go: Ethnic ingroup Bias in Judicial Decisions—Evidence from a Randomized Natural Experiment." *Journal of Empirical Legal Studies* 7(3): 403–28.

Gruhl, John, Cassia Spohn, and Susan Welch. 1981. "Women as Policy Makers: The Case of Trial Judges." *American Journal of Political Science* 25(2): 308–22.

Gryski, Gerard S., Eleanor C. Main, and William J. Dixon. 1986. "Models of State High Court Decision Making in Sex Discrimination Cases." *Journal of Politics* 48(1): 143–55.

Haj, Yahia, Muhammad, M. Rahav, and M. Teichman. 1994. "The Lower Court and Minorities in Israel." *Plilim: Israel Journal of Criminal Justice* 4: 157–74.

Hassin, Yael, and Michael Kremnitzer. 1993. "The Offence of Breaking and Entering—Variables Affecting Consistency in Sentencing." *Mishpatim* 22: 571–85.

Hilbink, Lisa. 2007. *Judges beyond Politics in Democracy and Dictatorship: Lessons from Chile.* New York: Cambridge University Press.

Hirano, Keisuke, and Guido W. Imbens. 2001. "Estimation of Causal Effects Using Propensity Score Weighting: An Application to Data on Right Heart Catheterization." *Health Services and Outcomes Research Methodology* 2(3–4): 259–278.

Holmes, Malcolm D., Harmon M. Hosch, Howard C. Daudistel, Dolores A. Perez, and Joseph B. Graves. 1993. "Judges' Ethnicity and Minority Sentencing: Evidence Concerning Hispanics." *Social Science Quarterly* 93(74): 496–506.

Honaker, James, Gary King, and Matthew Blackwell. 2011. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45(7): 1–47.

Ifill, Sherrilyn A. 2000. "Racial Diversity on the Bench: Beyond Role Models and Public Confidence." *Washington & Lee Law Review* 57: 405–95.

Johnson, Brian D. 2006. "Multilevel Context of Criminal Sentencing: Integrating Judge- and County-Level Influences." *Criminology* 44(2): 259–98.

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

Khaṭṭāb, Nabīl Mūsá, and Sami Miaari. 2013. *Palestinians in the Israeli Labour Market: A Multi-Disciplinary Approach.* New York, NY: Palgrave Macmillan.

Korn, Alina. 2003. "Rates of Incarceration and Main Trends in Israeli Prisons." *Criminology and Criminal Justice* 3(1): 29–55.

Kretzmer, David. 1990. *The Legal Status of the Arabs in Israel.* Boulder, CO: Westview Press.

Lerner, Hanna. 2011. *Making Constitutions in Deeply Divided Societies.* Cambridge: Cambridge University Press.

Manski, Charles F. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14(3): 115–36.

Martin, Elaine. 1990. "Men and Women on the Bench: Vive la Difference?" *Judicature* 73(4): 204–8.

McEvoy, Joanne, and Brendan O'Leary. 2013. *Power Sharing in Deeply Divided Places: National and Ethnic Conflict in the Twenty-First Century.* Philadelphia: University of Pennsylvania Press.

Mendelson, Wallace. 1964. "The Untroubled World of Jurimetrics." *Journal of Politics* 26(4): 914–22.

Nagel, Stuart S. 1962. "Testing Relations between Judicial Characteristics and Judicial decision making." *Western Political Quarterly* 15(3): 425–37.

Parsons, Christopher A., Johan Sulaeman, Michael C. Yates, and Daniel S. Hamermesh. 2011. "Strike Three: Discrimination, Incentives, and Evaluation." *American Economic Review* 101(4): 1410–35.

Peleg, Ilan, and Dov Waxman. 2011. *Israel's Palestinians: The Conflict Within.* Cambridge: Cambridge University Press.

Peresie, Jennifer L. 2005. "Female Judges Matter: Gender and Collegial Decisionmaking in the Federal Appellate Courts." *Yale Law Journal* 114(7): 1759–90.

Posner, Richard A. 2008. *How Judges Think.* Cambridge, MA: Harvard University Press.

Price, Joseph, and Justin Wolfers. 2010. "Racial Discrimination among NBA Referees." *Quarterly Journal of Economics* 125(4): 1859–87.

Rachlinski, Jeffrey J., Sheri Lynn Johnson, Andrew J. Wistrich, and Chris. Guthrie. 2009. "Does Unconscious Racial Bias Affect Trial Judges?" *Notre Dame Law Review* 84(3): 1195–1246.

Rattner, Arye, and Gideon Fishman. 1998. *Justice for All?: Jews and Arabs in the Israeli Criminal Justice System.* Westport, CT: Praeger.

Ridgeway, Greg, Dan McCaffrey, Andrew Morral, Beth Ann Griffin, and Lane Burgette. 2013. *twang: Toolkit for Weighting and Analysis of Nonequivalent Groups.* R package version 1.3-11.

Schanzenbach, Max. 2005. "Racial and Sex Disparities in Prison Sentences: The Effect of District-Level Judicial Demographics." *Journal of Legal Studies* 34: 57–92.

Shayo, Moses, and Asaf Zussman. 2011. "Judicial Ingroup Bias in the Shadow of Terrorism." *Quarterly Journal of Economics* 126(3): 1447–84.

Shotts, Kenneth W. 2003. "Does Racial Redistricting Cause Conservative Policy Outcomes? Policy Preferences of Southern Representatives in the 1980s and 1990s." *Journal of Politics* 65(1): 216–26.

Sommers, Samuel. 2006. "On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations." *Journal of Personality and Social Psychology* 90(4): 597–612.

Spohn, Cassia. 1990. "The Sentencing Decisions of Black and White Judges: Expected and Unexpected Similarities." *Law & Society Review* 24(5): 1197–1216.

Spohn, Cassia. 2009. *How Do Judges Decide? The Search for Fairness and Justice in Punishment.* 2nd ed. Los Angeles: Sage.

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.

Tajfel, Henry, and John Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, ed. William G. Austin and Stephen Worchel. Monterey, CA: Brooks/Cole, 33–47.

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

Weinshall-Margel, Keren. 2011. "Attitudinal and Neo-Institutional Models of Supreme Court Decision Making: An Empirical and Comparative Perspective from Israel." *Journal of Empirical Legal Studies* 8(3): 556–86.

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

Zamir, Yitzhak, Amnon Rubinstein, and Yoram Guy-Ron. 2001. "Committee for Evaluating Judges Selection Procedure." Technical report, Ministry of Justice, Israel.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Introduction
Robustness Check I: Court-Specific Analyses
Robustness Check II: Parsimonious Models
Robustness Check III: Interactions
Robustness Check IV: Transforming the Response
Robustness Check V: Propensity Score Weighting
Text Analysis
Descriptive Statistics
Technical supplement: marginal effect calculation