Contagious Animosity in the Field: Evidence from the Federal Criminal Justice System

Brendon McConnell, *University of Southampton*

Imran Rasul, *University College London*

We investigate whether increased animosity toward Muslims after 9/11 had spillover effects on Black and Hispanic individuals in the federal criminal justice system. Using linked administrative data tracking defendants from arrest through sentencing, we find that after 9/11, sentence and presentence outcomes for Hispanic defendants significantly worsened. Outcomes for Black defendants were unchanged. The findings are consistent with judges and prosecutors displaying social preferences characterized by contagious animosity from Muslims to Hispanics. Our findings provide among the first field evidence of contagious animosity, indicating that social preferences across out-groups are interlinked and malleable.

We gratefully acknowledge financial support from the Dr. Theo and Friedl Schöeller Research Center for Business and Society and the Economic and Social Research Council (ESRC) Centre for the Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies (grant RES-544-28-5001). We thank Oriana Bandiera, Patrick Bayer, Daniel Bennett, Marianne Bertrand, Pietro Biroli, Dan Black, David Card, Kerwin Charles, Steve Cicala, Gordon Dahl, Brad DeLong, Ben Faber, Rick Hornbeck, Randi Hjalmarsson, Emir Kamenica, Neale Mahoney, Alan Manning, Olivier Marie, Ioana Marinescu, Michael Mueller-Smith, Aviv Nevo, Emily Owens, Daniele Paserman, Steve Pischke, Steven Raphael, Jesse Rothstein, Anna Sandberg, Johannes Schmeider, Robert Topel, and numerous seminar and conference participants for valuable comments. All errors remain our own. Contact the corresponding author, Imran Rasul, at i.rasul@ucl.ac.uk. Information concerning access to the data used in this paper is available as supplemental material online.

[*Journal of Labor Economics*, 2021, vol. 39, no. 3]
 0734-306X/2021/3903-0005$10.00
Submitted April 19, 2018; Accepted August 3, 2020; Electronically published May 11, 2021

I. Introduction

Minority men are far more likely to come into contact with the federal criminal justice system (CJS) than White men, and decades of research have shown sentencing outcomes vary by race and ethnicity. The challenge in interpreting such sentencing differentials lies in establishing whether they are driven by unobserved heterogeneity correlated to defendant race/ethnicity or whether they reflect discrimination. The question is of fundamental importance given that equality before the law is a cornerstone of any judicial system and because it is difficult to know whether and how to reduce sentencing disparities if their underlying causes remain unknown.

We advance this literature using three novel pillars of analysis to identify and measure the decisions of judges and prosecutors that determine differential outcomes by race/ethnicity. The building blocks underlying our analysis are modifying the notion of in-group and out-group bias in societies comprising multiple groups/identities, using a novel research design built around this notion, and exploiting linked administrative records tracking defendants through all stages of the federal CJS.

A vast literature examines the biological and evolutionary roots of in-group bias (Tajfel et al. 1971). Individuals are assumed to have some social preference over the payoffs to their in-group and their out-group, where they favor their in-group more strongly. As with individual preferences, the standard view is that such social preferences are stable and immutable.[1] However, there has been increasing attention on alternative formulations that suggest that such social preferences are malleable. A nascent body of laboratory evidence shows that agents can display contagious altruism: under this view, positive altruism toward an out-group fosters altruism toward the in-group. A second scenario is one of parochial altruism: under this view, greater rivalry between groups fosters more cooperation within the in-group.[2]

We apply these notions to US society, where individuals can have one of many identities. There is thus one in-group and multiple out-groups, so social preferences are defined over all of these groups. We then ask whether increased animosity toward one out-group drives social preferences toward another out-group. The answer is no if social preferences across out-groups are independent. On the other hand, there can be "contagious animosity" across

[1] Social psychologists have documented dimensions such as race, ethnicity, religiosity, and political affiliation as all being salient across contexts in driving in-group biases. In economics, in-group biases have been studied in laboratory settings and shown to emerge even in artificially created groups (Shayo 2009; Bertrand and Duflo 2017).

[2] Contagious altruism has been documented in laboratory settings (Fowler and Christakis 2010; Suri and Watts 2011; Jordan et al. 2013). The idea of parochial altruism goes back to Darwin and has gained traction in economics, anthropology, political science, and psychology (Alexander 1987; Boyd et al. 2003; Eifert, Miguel, and Posner 2010). Much of this relies on self-reports or lab-in-field studies in post-conflict societies (Bauer et al. 2016).

out-groups, such that hostility toward one out-group drives hostility toward others. Alternatively, there might be "parochial animosity," such that hostility toward one out-group increases altruism toward other out-groups. While the study of in-group–out-group biases goes back decades, to the best of our knowledge there has been little examination of spillover effects across out-groups (Bertrand and Duflo 2017). The notion is important because it implies that out-group biases are malleable and that antidiscrimination policies against one out-group can have positive or negative externalities on other out-groups.

We use the ideas of contagious/parochial animosity to construct a research design to examine racial/ethnic sentencing differentials in the federal CJS, a high-stakes and professional economic environment. This is a setting in which defendants are of multiple identities (by race, ethnicity, citizenship, etc.) and the vast majority of federal judges and prosecutors during our study period are White, so we view them as the in-group. We consider 9/11 as an exogenously timed event that heightened the salience of insider-outsider differences in US society and, specifically, that increased animosity toward Muslims (Human Rights Watch 2002; Davis 2007; Woods 2011). We use this exogenously timed shock toward one out-group to measure spillovers on sentencing outcomes in the CJS for other out-groups, namely, for Black and Hispanic defendants.

A priori, not all out-groups would be equally impacted through spillovers induced by the structure of social preferences. In particular, there are reasons why Hispanic defendants are closer to Muslims in social construct than other out-groups. Drawing on work in sociology, we provide a detailed account of how Islamophobia and immigration have become gradually intertwined in American consciousness since the mid-1990s but were most forcefully framed together in the aftermath of 9/11 (Romero and Zarrugh 2018). Three channels are identified linking Islamophobia and Hispanics: (i) political rhetoric, (ii) policy framing, and (iii) restructured institutions.

We examine the impact of 9/11 on sentencing gaps across races/ethnicities using the Federal Justice Statistics Resource Center (FJSRC) data combined with the Monitoring of Federal Criminal Sentences (MFCS) data set (USSC 1999–2003). This covers the universe of all male defendants up for sentencing from 1998 to 2003, so either side of 9/11 and totaling 230,000 federal criminal cases. It is nationally representative, covering cases from all 90 mainland US districts, defendants of all ages, and all types of criminal offense. Such large and representative samples allow for both Black-White and Hispanic-White differentials to be studied. Moreover, the FJSRC comprises four linked administrative data sources covering the time from a defendant's initial arrest and offense charge and all subsequent stages of their processing through the federal CJS. This linked administrative data set thus allows presentencing differential treatment arising from the behavior of prosecutors or legal counsel to be studied alongside the behavior of judges at sentencing. Furthermore, it enables us to pin down whether judges and prosecutors display similar kinds of social preference structures across out-groups and to address longstanding challenges for empirical work on the CJS that is typically based on sentencing data only (Klepper, Nagin, and Tierney 1983).

The FJSRC-MFCS data do not allow direct impacts of 9/11 on Muslim defendants to be studied because they contain no identifier for religion. Even if they did, very few defendants of Muslim origin would be expected in the federal CJS in our study period.

To isolate the impact 9/11 had on sentencing outcomes, we compare (i) defendants who committed their last offense before 9/11 and were sentenced before 9/11 with (ii) defendants who also committed their last offense before 9/11 but were sentenced after 9/11. We construct a second difference in outcomes across race/ethnicity to estimate a difference-in-differences impact of 9/11 on sentencing outcomes. We base our sample on a ±180-day sentencing window around September 11, 2001, where all defendants have committed their offense before 9/11 and hence entered stage 1 of the federal CJS timeline in figure 1, but some were sufficiently far advanced along the timeline so as to come up for sentencing before 9/11, while others had only just entered the timeline before 9/11 and so ended up being sentenced after 9/11.

The period we study is when sentencing guidelines are in place. These guidelines provide for determinate sentencing, mapping combinations of the severity of the offense and the defendant's criminal history into a sentencing range. Table A1 (tables A1–A14 are available online) shows the full set of guideline cells. The guidelines do, however, allow judges' discretion to "downward depart" from the recommended guideline cell and so move in a northerly direction in table A1. This is the primary outcome of interest when studying judicial decision-making and is an important margin to consider. For example, Mustard (2001) documents that 55% of the Black-White sentencing differential is attributable to differences in downward departure.

Our core results are as follows. We first confirm that relative to Whites, Blacks and Hispanics sentenced before 9/11 receive significantly longer prison sentences. For Hispanics sentenced after 9/11, sentencing differentials become further exacerbated through a specific channel: they become 13.5% less likely to receive a downward departure than Whites. The implied increase in sentence length for Hispanics is .736 months, corresponding to 18% of the conditional pre-9/11 differential in sentence length. Placing a monetary value on this increased incarceration suggests that the spillover effects from heightened animosity toward Muslims after 9/11 led to an increase of $1,547 in incarceration costs per Hispanic defendant. This maps to a large increase in total costs for the federal CJS given that the modal defendant in the study period is Hispanic.

We further develop an approach to identify the marginal defendants most likely to be impacted by changes in judges' propensity to downward depart. We find that among marginal defendants, 9/11 led to a increased Hispanic-White sentence differential of just over 2 months, corresponding to 50% of the conditional pre-9/11 differential in sentence length. The magnitude of

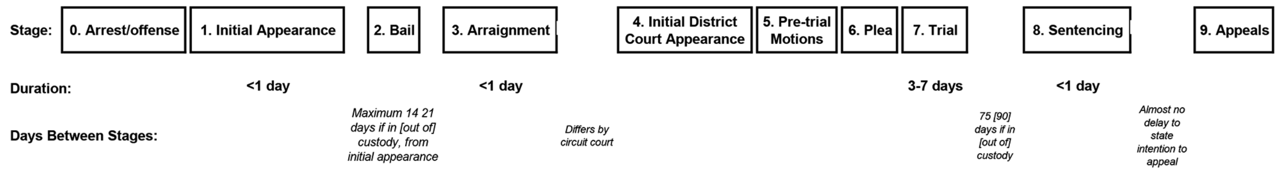

Fig. 1.—Federal criminal justice system timeline. We use Federal Justice Statistics Resource Center (FJSRC) data and the Monitoring of Federal Criminal Sentences (MFCS) data set for our analysis. The FJSRC data comprise information gathered from four linked administrative data sources, and we replace the fourth-stage US Sentencing Commission (USSC) data, which covers sentencing stage 8, with the MFCS data (because it is superior to the USSC data, which is part of the FJSRC). As described in Rehavi and Starr (2014), the linked data sets are (i) US Marshals Service data, which cover the arrest/offense stage (stage 0) and include all persons arrested by federal law enforcement agencies, persons arrested by local officials and then transferred to federal custody, and persons who avoid arrest by self-surrendering; (ii) Executive Office for US Attorneys data, which cover initial appearance through arraignment (stages 1–3; these data come from the internal case database used by federal prosecutors and cover every case in which any prosecutor at a US attorney's office opens a file); (iii) Administrative Office of the US Courts data, which cover initial district court appearances through trial (stages 4–7; these data originate from federal courts and contain information on all criminal cases heard by federal district judges and any nonpetty charge handled by a federal magistrate judge); and (iv) the MFCS data set, which we use at stage 8.

this is comparable to sentencing differentials across groups that opened up after sentencing guidelines were abolished altogether in 2005 (Yang 2015).

Black-White sentencing differentials around 9/11 are unaffected along all sentencing margins, and as far as the data allow, we find the post-9/11 impacts to be statistically similar for Hispanic citizens and Hispanic noncitizens. Overall, the results are consistent with judges displaying contagious animosity from Muslims to Hispanics, while their social preferences are independent between Muslim and Black defendants, and we find no evidence that 9/11 leads to greater altruism within the majority in-group.

To underpin a causal interpretation, we provide evidence in support of the identifying assumptions underlying our research design. We first show that the time a defendant spends in the CJS between when their last offense is committed and when they come up for sentencing is not impacted by 9/11. Hence, there is no evidence of resequencing of cases by race/ethnicity after 9/11. Second, using data from other years to construct placebo 9/11 impacts, we show that there are no natural race/ethnicity-time effects in sentencing differentials that occur around 9/11 each year. Third, we show that the estimates are robust to selection on unobservables, ruling out plausible changes in Hispanic-specific unobservable factors after 9/11 that could drive the main finding.

Our data and research design allow us to probe beyond judges' sentencing decisions. As has long been recognized (Klepper, Nagin, and Tierney 1983), a range of legal actors beyond judges are involved in the timeline of federal criminal cases, and their behaviors can lead to differential treatment presentencing, which might not be detected in sentencing differentials. These concerns are heightened when sentencing guidelines are in place, as these restrict the discretion of judges and might increase the power of prosecutors, especially in a system characterized by plea bargaining (Starr and Rehavi 2013). We use the linked administrative data and our research design to move our 9/11 window to earlier stages of the case timeline in figure 1, where key decisions by prosecutors are being made.

As with judges, the results on prosecutors' decisions are consistent with them displaying contagious animosity from Muslims to Hispanics and their social preferences being independent between Muslim and Black defendants. More precisely, Hispanic defendants initially charged after 9/11 are 7.5 percentage points more likely to receive an initial offense that carries a statutory minimum, and their statutory minimum sentence is 10.7 months longer. These impacts correspond to (i) 60% of the pre-9/11 Hispanic-White gap in the likelihood of an initial offense charge with a mandatory minimum and (ii) 77% of the pre-9/11 Hispanic-White gap in the statutory minimum sentence length. Indeed, these causal responses to 9/11 lead the Hispanic-White differential on each margin to become as large as the pre-9/11 Black-White differential.[3]

[3] On prosecutorial biases, Rehavi and Starr (2014) use related linked administrative data from the federal CJS to show that prosecutor's initial offense charges account for

Having established a causal spillover of 9/11 on Hispanic outcomes in the federal CJS, our final set of results probe the data to narrow the interpretation of these widening Hispanic-White differentials. As best as the data allow, we explore whether the results can be explained through statistical discrimination (say through higher expected recidivism rates of Hispanics after 9/11).

We first present a Juhn-Murphy-Pierce (1993) decomposition of sentencing differentials between those who come up for sentencing after 9/11, where Hispanics are significantly less likely to receive a downward departure from judges. The decomposition shows that only negligible amounts of the unconditional difference-in-differences in outcome can be attributed to either differences in their observables relative to Whites or the sentencing penalties of such observables. This helps to rule out explanations for the increased Hispanic-White differential based on the harshness with which certain offense types are dealt with after 9/11; offender characteristics, including those that might perhaps closely predict recidivism, such as the guideline cell they are assigned to; or explanations related to effort or allocation of legal counsel to defendants after 9/11. Overall, the decomposition suggests that explanations for why Hispanic-White sentencing differentials worsen after 9/11 based on statistical discrimination do not easily fit the evidence.

Second, we analyze how judge characteristics correlate to the estimated Hispanic-White sentencing differential. We code characteristics of federal judges by district court, sourced from the *Biographical Directory of Federal Judges*. We document that in districts with a higher proportion of Hispanic federal judges, the post-9/11 Hispanic-White sentencing differential for downward departures is significantly reduced. The fact that judge ethnicity correlates to the Hispanic-White sentencing differential is again prima facie evidence against the results being explained by statistical discrimination: if so, then all judges, irrespective of their own characteristics, should use race/ethnicity as a marker for unobservable traits in determining sentencing outcomes. This is in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using data on police arrests or on individual judges (Anwar and Fang 2006; Park 2017).

Both strategies suggest that 9/11 had spillover effects on Hispanics through decisions made by judges, with them having social preferences displaying contagious animosity from Muslim to Hispanic out-groups but independence between Muslim and Black out-groups. Our analysis contributes to two long-standing literatures: on in-group and out-group biases as drivers of human behavior and on sentencing differentials in the CJS.

We provide among the first field evidence based on a quasi-experimental research design of the existence of contagious animosity. We do so in the high-stakes and professional environment of the federal CJS. Earlier work

half the Black-White sentencing gap. They do so for the period 2006–8, after sentencing guidelines have been abolished.

on sentencing differentials in other parts of the CJS has explicitly or implicitly framed the issue in terms of in-group and out-group biases (Bushway and Piehl 2001; Shayo and Zussman 2011; Abrams, Bertrand, and Mullainathan 2012; Anwar, Bayer, and Hjalmarsson 2012; Rehavi and Starr 2014). By allowing for multiple out-groups and developing the notion of contagious/parochial altruism, our work has the important implication that in multigroup societies, effective antidiscrimination policies targeting one group can have positive externalities on other minority groups. Our analysis also helps address an appeal made in recent overviews of the economics of discrimination literature on the need to better bridge to the psychology literature on the origins of discriminatory behavior (Charles and Guryan 2011; Bertrand and Duflo 2017).

The literature has studied three sources of racial/ethnic sentencing differentials (Fischman and Schanzenbach 2012): (i) judicial bias, (ii) prosecutorial bias, and (iii) sentencing policies. The linked administrative data we use provides insights into the first two dimensions. We advance the literature by pinpointing the separate roles that judges and prosecutors have in driving the differential treatment of Hispanics in the federal CJS after 9/11 and explaining the behavior of both through the structure of their social preferences across multiple out-groups.

The paper is organized as follows. Section II describes the federal CJS, sentencing guidelines, and administrative data. Section III presents motivating evidence on long-standing pre-9/11 sentiments against Hispanics and then builds an evidence base to argue how 9/11, Islamophobia, and immigration issues all became interlinked in the aftermath of 9/11. Sections IV and V present our core findings on sentencing differentials, as driven by the decision-making of judges and prosecutors, respectively. Section VI narrows the interpretation of increased sentencing differentials after 9/11 using decomposition analysis and judge characteristics. Section VII concludes. The appendix (available online) contains further data details, robustness checks, and additional results.

II. The Federal Criminal Justice System

Criminal cases are filed in federal court if prosecuted by a federal agency or related to federal law. In 2000, the three most frequent criminal offenses were for drug trafficking (40%), immigration (22%), and fraud (9%). This is a high-stakes setting: cases heard in federal courts tend to be more serious than those heard in state courts. Eighty-eight percent (75%) of those convicted in federal (state) court receive a custodial sentence, with the mean sentence being 67 (48) months in federal (state) court.[4]

[4] If both federal and state courts have jurisdiction over a criminal act, prosecutors make case-by-case decisions on which court the defendant will be tried in, although the presumption is that federal prosecutors hold greater sway in such decisions given the greater resources at their disposal (Jeffries and Gleeson 1995). The sorting of cases into systems is therefore an executive branch decision: judges and defense counsel

The primary legal actors determining outcomes in federal criminal cases are judges, prosecutors, and legal counsel. Federal judges are presidential nominees, confirmed by Congress, and life appointees. Prosecution in each of the 94 US district courts is the responsibility of the US attorney for that district, who is also a presidential appointee reporting directly to the attorney general. There are around seven federal judges per district, so close to 700 in total. They are among the most senior judges and a priori might be considered among those least susceptible to displaying contagious/parochial animosity across out-groups.

In 47% of federal criminal cases, legal counsel is court appointed. Federal public defenders operate in 32% of cases, and 21% of defendants retain private counsel. This differs from state court cases, where 68% of defendants have a public defender. Finally, jury trials in federal courts occur only if a defendant pleads not guilty. In the federal CJS this is rare: 96% of defendants plead guilty before they reach trial. By pleading guilty, the individual is convicted, and only their sentence remains to be determined. Guilty pleas can be taken into account at sentencing, and such pleas can be Pareto improving for risk-averse defendants and prosecutors. By pleading guilty, defendants give up the right to appeal except in capital cases (less than 0.1% of cases).

A. Timeline

Figure 1 shows the timeline for federal criminal cases, as covered in the FJSRC data. Table A2 further details each stage. The first stage a defendant faces after having been arrested and formally charged with a federal offense (stage 0) is their initial court appearance, where their defense counsel is assigned (stage 1). Bail is then determined (stage 2) and initial charges are filed by prosecutors during arraignment (stage 3), leading to the defendant's initial district court appearance (stage 4), where they find out which judge they have been assigned to. Pretrial motions take place at stage 5, to determine what evidence can be used in trial. The defendant can then offer a plea (stage 6), where 96% plead guilty, and defendant cooperation can be rewarded by prosecutors. The trial represents stage 7, and sentencing occurs at stage 8. In rare cases where a defendant pleads not guilty or for capital cases, they retain the right to appeal (stage 9).

Two other aspects of the timeline are of note. First, a magistrate judge handles the first stages of a defendant's passage through the CJS. At arraignment, the magistrate will issue a scheduling order and determine which district

have no formal role. The difference-in-differences research design we use to estimate sentencing differentials eliminates cross-sectional differences between defendants, by race, being sent to trial in the federal system. Glaeser, Kessler, and Piehl (2000) provide a theoretical and empirical analysis of the sorting of cases into state and federal systems. The difference in severity across courts is not driven by the composition of offenses: within offense type there is considerably harsher sentencing in federal courts, reflecting the greater seriousness of such crimes.

court judge will actually preside over the case. With the exception of pretrial motions, hearings that are heard by the magistrate, the district court judge presides over the rest of the case (stage 6 onward). Second, the recommended guideline cell is determined between trial and sentencing (stages 7 and 8): this is when the presentence report is drafted by the (neutral) Probation Office, the defendant's legal counsel, and prosecutors. A fortnight before sentencing, the final presentence report is presented to the judge. This describes the defendant's background and offense (including the impact on the victim). It reports a determined criminal history score and the offense severity and thus the recommended guideline cell.

We first focus on sentencing (stage 8). As 96% of defendants are already convicted, only their punishment is to be determined. This is where judges exercise discretion. Multiple legal actors are involved at earlier stages; their behaviors can lead to differential treatment of defendants presentencing, and the presence of biases earlier in the timeline might not be detected in judicial sentencing differentials. In section V we exploit the linked administrative data to consider earlier stages to pinpoint how prosecutors drive sentencing differentials, including the initial offense charges of prosecutors, which have been shown to play an important role in Black-White sentencing gaps (Rehavi and Starr 2014).

B. Linked Administrative Data

The FJSRC data set comprises four linked administrative data sources covering the arrest/offense stage before an individual enters the federal CJS (stage 0) and all subsequent stages shown in figure 1. For sentencing stage 8, we use the MFCS data (which can be linked to earlier data sets in the FJSRC).[5] We focus on male defendants. Our sample covers 230,000 federal criminal cases up for sentencing from October 1998 to September 2003 across nearly all US districts. The appendix provides further data details. To estimate Black-White and Hispanic-White sentencing differentials, we use two variables available at sentencing stage 8. In the first, defendants are coded as Hispanic (41%) or non-Hispanic (59%). A separate race code then identifies defendants as white race (71%), black race (29%), or other race (<0.1%). We code Whites as white-race non-Hispanic, Blacks as black-race non-Hispanic, and Hispanics as white- or black-race Hispanics. This implies that 31% of defendants are White, 26% are Black, and 43% are Hispanic.

The data detail defendant demographics, including age, highest education level, marital status, citizenship, and number of dependents. Legal controls include the type of defense counsel, other presentence variables (such as whether the defendant is in custody), and the federal court district, and we

[5] As explained in the appendix, the MFCS data are superior to the US Sentencing Commission (USSC) data in the FJSRC (even though it also originates from the USSC) because they contain exact sentence dates and dates of last offense.

use offense details to classify 31 offense types.[6] Most importantly, the data record the guideline cell recommended to the judge in the presentence report. This effectively proxies all case-specific factors the prosecution and legal counsel deem judges should factor into sentencing. However, the data do not identify the cell the defendant was then placed into if downward departed: we observe only the sentence length, which as figure A1 (figs. A1–A4 are available online) makes clear might correspond to many different cells. We later detail the algorithm we use to provide an indication of the number of cells moved conditional on being downward departed.

A concern when studying sentencing outcomes is that there can be selection of defendants such that cases reaching sentencing might not be representative of the original population of charged defendants (Klepper, Nagin, and Tierney 1983). As the FJSRC-MFCS data comprises linked administrative sets covering arrest/offense (stage 0) through sentencing (stage 8), we can estimate dyadic linkage rates for criminal cases across stages of the timeline. In the appendix, we show that these linkage rates are similar by race/ethnicity and by offense type. The difference-in-differences research design we use to estimate sentencing differentials eliminates cross-sectional differences between defendants of different race/ethnicity (such as in linkage rates).

C. Federal Sentencing Guidelines

Federal sentencing guidelines were introduced in the Sentencing Reform Act of 1984 by the USSC. The goal was to alleviate sentencing disparities through determinate sentencing, limiting the discretion judges had over sentencing. Parole boards were also abolished, such that actual incarceration length became a fixed threshold of 85% of determined sentences.

The sentencing guidelines are based on (i) the severity of the offense and (ii) the defendant's criminal history. To run through a stylized example, an individual who commits a robbery is allocated a base level of 20 points. If a gun is involved, an additional 5 points are awarded (if the individual had been a minimal participant in the robbery, 4 points would have been deducted). If the individual was found to be in obstruction of justice, an additional 2 points are awarded. Hence, in this case the final score of the defendant on offense severity would be 23 points. There are six criminal history categories, each associated with a range of criminal history points. Criminal history points are based on each prior sentence of imprisonment (and vary with the length of that

[6] These include kidnaping/hostage taking; sexual abuse; assault; bank robbery (including arson); drugs: trafficking; drugs: communication; drugs: simple possession; firearms: use (including burglary/breaking and auto theft); larceny; fraud; embezzlement; forgery/counterfeiting; bribery; tax offenses; money laundering; racketeering (including gambling/lottery); civil rights offenses; immigration; pornography/prostitution; offenses in prisons; environmental; national defense offenses; antitrust violations; food and drug offenses; traffic violations; and other smaller categories.

earlier imprisonment), whether the offense was committed while under parole/release, and so on. Suppose the individual in the example given above was assessed to have 7 criminal history points. The sentencing guidelines then stipulate that the person should be sentenced in the range of 70–87 months.

Table A1 shows the full set of guideline cells, mapping each combination of offense severity (1–43) and criminal history (1–13, grouped into six bins) into a sentencing range. There are $43 \times 6 = 258$ guideline cells. These include those in zone A, where the guidelines include zero sentence length, and those in zone D, where the guidelines impose a life sentence.

Between trial/conviction and sentencing (stages 7 and 8), the presentence report is drafted by prosecutors, legal counsel, and an independent probation officer. This recommends a guideline cell. However, the guidelines still provide judges discretion to downward depart from the recommended guideline cell and move in a northerly direction in the guideline cell table A1. A judge can do so if they find mitigating circumstances of a kind not adequately taken into consideration by the USSC in formulating the sentencing guidelines. These circumstances include diminished capacity or rehabilitation after the offense but before sentencing, family responsibilities, or prior good works. Downward departures may also be warranted if "information indicates that the defendant's criminal history category substantially over-represents the seriousness of the defendant's criminal history or the likelihood that the defendant will commit other crimes." Judges are required to provide written explanations for their reason(s) for downward departing.

In our sample, judges grant downward departure in 17% of cases. This results in a sentence below the original guideline range, but they still lead to a custodial sentence in 90% of cases. Upward departures occur in less than 1% of cases. Judge-initiated downward departures are the key sentencing outcome to consider because (i) such decisions are cleanly attributable to judges and (ii) they are associated with reductions in sentence length.

The null hypothesis for our analysis is based on the USSC sentencing guidelines that state that "race, sex, national creed, religion and socioeconomic status" are factors that "are not relevant in the determination of a sentence" (§5H1.10 of the sentencing guidelines).[7]

III. Descriptives, 9/11, and Research Design

A. Pre-9/11 Sentencing Differentials

We examine pre-9/11 sentencing differentials along two margins of judicial decision-making: (i) if a downward departure is granted and (ii) the sentence length (in months).

[7] The guideline cells were in operation until 2005. The Supreme Court's 2005 decision in *United States v. Booker* found the mandatory application of guidelines to be unconstitutional. The guidelines are now considered advisory.

Columns 1 and 3 in table 1 show unconditional differentials by race/ethnicity for each outcome. Black-White and Hispanic-White differentials are of statistical and economic significance. We next examine whether these differentials are robust to conditioning on a rich set of covariates including the demographic characteristics of the defendant described earlier (X_i), the type of legal counsel (L_i), offense type (OFF_{if}), the guideline cell they are assigned to in the presentence report (G_{ig}), dummies for the federal court district in which the case is considered (D_{id}), and dummies for fiscal year t, π_t. A key advantage of using the MFCS data for sentencing outcomes is that we can nonparametrically condition on the full set of guideline cells. This effectively proxies all case-specific factors that prosecutors and legal counsel deem judges should factor into their sentencing decision (such as whether a gun was used in the crime, the quality of drugs involved in drug offenses, etc.). Such factors would typically be unobserved by the econometrician.

Table 1
Pre-9/11 Sentencing Differentials in Judge's Decisions

	Downward Departure		Sentence Length	
	Unconditional (1)	Conditional (2)	Unconditional (3)	Conditional (4)
Black	−.047***	−.008**	42.2***	3.88***
	(.015)	(.004)	(2.57)	(.523)
Hispanic	.133***	.010	1.72	4.08***
	(.050)	(.011)	(3.71)	(.540)
Sentencing outcome for Whites	.125		40.5	
Offender, legal, and district controls	No	Yes	No	Yes
Offense type codes	No	Final	No	Final
Guideline cells	No	Yes	No	Yes
p-value: Black = Hispanic	.002	.037	.000	.736
Adjusted R^2	.044	.242	.064	.743
Observations	130,895	130,895	130,895	130,895

NOTE.—Ordinary least squares regression estimates are shown in all columns except 3 and 4, where a negative binomial specification is estimated. Standard errors clustered by district are reported in parentheses. The pre-9/11 sample of 130,895 federal cases is used (those that come up for sentencing from October 1, 1998, to September 10, 2001). The dependent variable in cols. 1 and 2 is a dummy for whether the case receives a downward departure. The dependent variable in cols. 3 and 4 is the sentence length (in months) including zero. In cols. 1 and 3 we condition only on defendant group (White, Black, Hispanic). In cols. 2 and 4 the following additional controls are included: fiscal year dummies, offender characteristics (dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this nonmissing-age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a nonmissing-dependents dummy), legal controls (a dummy for whether information on the defense counsel is missing and a nonmissing dummy interacted with the type of defense counsel [privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements]), the primary offense type, the guideline cell, and federal district dummies. The p-value in each column is for the null that the coefficients on the Black and Hispanic dummy are equal against a two-sided alternative.

** Significant at 5%.
*** Significant at 1%.

Columns 2 and 4 show that conditioning on covariates, there are large changes in the Black- and Hispanic-dummy coefficient estimates on each margin. This is expected given that defendants in each group differ on observables. However, even conditional on covariates including the recommended guideline cell, we see that statistically significant Black-White and Hispanic-White sentencing differentials remain. For example, Black and Hispanic defendants have significantly longer sentence lengths. A natural benchmark we use for the later analysis on any spillover impacts of 9/11 on out-groups is the pre-9/11 conditional sentencing gap, which is around 4 months for both out-groups relative to Whites, or around 10% of the White sentence length.

B. Linking Muslim and Hispanic Out-Groups

We aim to understand whether judges and prosecutors display social preferences characterized by contagious or parochial animosity across out-groups. We do so by exploiting 9/11 as an exogenously timed increase in animosity toward one out-group: Muslims. The events of 9/11 certainly increased animosity toward Muslims (Human Rights Watch 2002; Davis 2007; Woods 2011) and reduced their rates of assimilation (Gould and Klor 2016). Not all out-groups would be impacted by any resulting contagious/parochial animosity, but there are reasons why Hispanics are closer to Muslims in social construct than other out-groups. To understand the link between 9/11 and Hispanics, we draw on work in sociology by Romero and Zarrugh (2018). They provide a detailed account of how Islamophobia and immigration have become gradually intertwined in American consciousness since the mid-1990s but were most forcefully framed together in the aftermath of 9/11. They build an evidence base for this thesis by analyzing government reports, media accounts, nongovernmental evaluations, statements by politicians, and other secondary sources. They argue that Islamophobia—or the extreme and irrational fear of Muslims and Islam—was deployed against Hispanics to garner political support and justify increased surveillance and immigration enforcement. Romero and Zarrugh (2018) identify three channels linking Islamophobia and Hispanics: (i) political rhetoric, (ii) policy, and (iii) institutions.

On political rhetoric, around 9/11 numerous politicians explicitly linked the events to immigration. Issues of security and threats to the nation were tied to immigration and specifically to the US-Mexico border. On policy, immigration and terrorism issues have slowly become intertwined since the 1995 Oklahoma bombings. Two prominent legislative acts linked immigration and terrorism before 9/11: the Illegal Immigration Reform and Responsibility Act and the Antiterrorism and Effective Death Penalty Act. Both became law in 1996, linking terrorism and immigration and broadening the set of federal criminal cases subject to deportation. After 9/11, the Patriot Act came into effect 45 days later, further increasing the link between terrorism and immigration through its near-exclusive focus on immigration offenses. On institutions, the formation of the Department of Homeland Security (DHS) represented the first time that terrorism and immigration agencies had been merged. The DHS merged 22 federal agencies, and as such the culture of the joint bureaucracy changed.

All three channels led to claims that "the war on terror quickly turned into the war on immigrants" (A. D. Romero, executive director, American Civil Liberties Union; Liptak 2003).

To provide quantitative evidence on the impacts on Hispanics in the immediate post-9/11 period, figure 2*A* shows time-series evidence from a Gallup poll on immigration: this highlights a marked shift against immigration after 9/11. Figure 2*B* shows vandalism victimization rates, by race/ethnicity. The data show a spike in Hispanics reporting being victims of vandalism after 9/11, with the growth rates in victimization rates only slowly returning back to trend. Other studies show that 9/11 worsened labor market outcomes for Hispanics (Orrenius and Zavodny 2009).[8]

Taken together, these rhetorical, policy, and institutional links among 9/11, immigration, and Hispanics leave open the possibility that outcomes for Hispanic defendants might be impacted in the aftermath of 9/11 if judges and prosecutors have social preferences across out-groups characterized by contagious/parochial animosity.

C. Research Design

To isolate the impact that 9/11 had on sentencing outcomes, we compare outcomes between (i) defendants who committed their last offense before 9/11 and were sentenced before 9/11 and (ii) defendants who also committed their last offense before 9/11 but were sentenced after 9/11. We construct a second difference in outcomes across race/ethnicity to estimate a difference-in-differences impact of 9/11 on criminal sentencing. Our working sample is based on a ±180-day sentencing window around September 11, 2001, where all defendants committed their offense before 9/11 and hence entered the federal CJS timeline in figure 1; however, some were sufficiently far advanced so as to come up for sentencing before 9/11, while others had only just entered the timeline before 9/11 and so ended up being sentenced after 9/11. To maintain comparability of both groups, we restrict the sample further so that for those defendants sentenced before 9/11 their last offense was committed at least 180 days before 9/11.[9]

[8] Legewie (2013) documents worsening attitudes toward immigrants in response to terrorist events in a range of countries; Hopkins (2010) uses panel data around 9/11 to show that it had a profound short-run impact on attitudes toward immigrants.

[9] We keep cases in which (i) guilty pleas are filed (which is so for 96% of defendants) and (ii) three or fewer offenses were committed, because for offenses that come up for sentencing from October 1, 2001, through September 30, 2002, we observe only the date of the first three offenses.

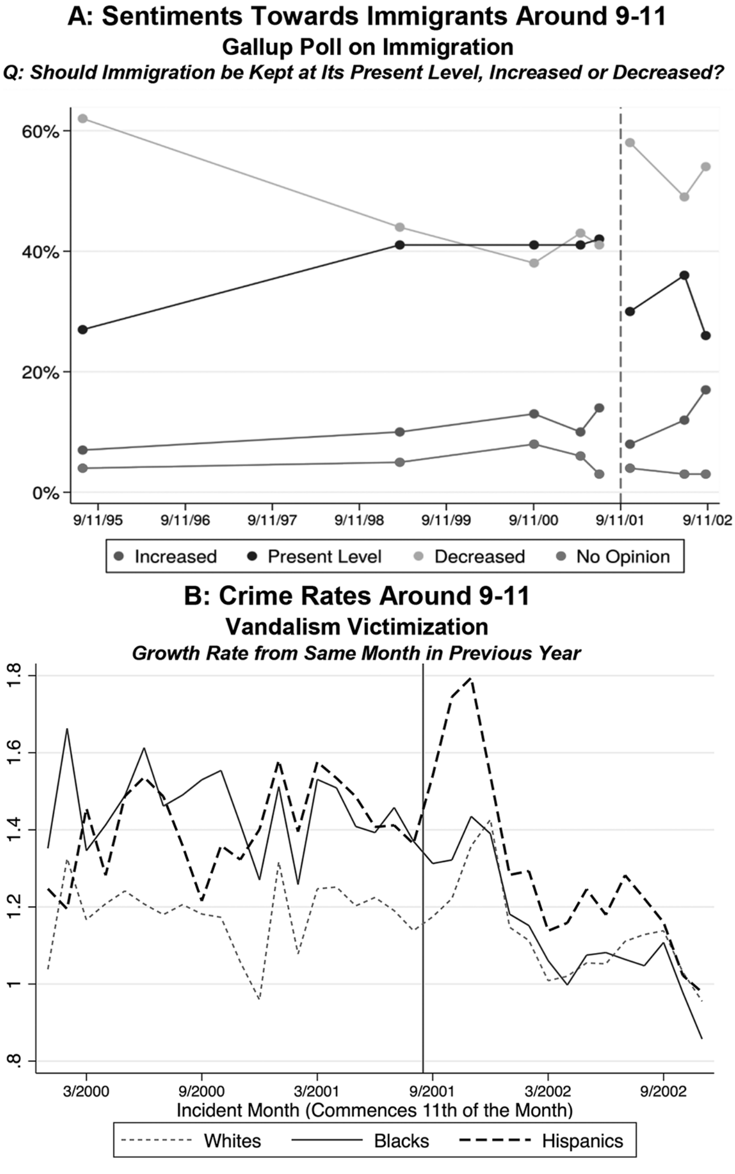

FIG. 2.—Pre- and post-9/11 sentiments. *A* is based on a Gallup poll that asks respondents, "Thinking more about immigration—that is, people who come from other countries to live here in the United States, in your view, should immigration be kept at its present level, increased or decreased?" The data were accessed via http://www.gallup.com/poll/1660/immigration.aspx. *B* is based on data from the National Incident-Based Reporting System Extract Files. The outcome variable is vandalism victimization. The data were collapsed to the month level, where month was constructed to start on the 11th in order to align with September 11, 2001. To account for seasonal

The working sample covers 40,228 cases: 32% of defendants are White, 27% are Black, and 41% are Hispanic. Table 2 shows the characteristics of each group of defendants, for cases up for sentencing before and after 9/11. The samples are well balanced on these defendant and legal characteristics, and the difference-in-differences in characteristics are nearly all not different from zero. Where there is imbalance, the magnitudes are small. Given that 9/11 was unanticipated, our evidence is based on a sample of defendants and offenses that are representative of caseloads in the federal CJS more broadly.

Figure 3 provides a graphical description of the research design by plotting histograms of the dates of sentencing and last offense for defendants. Focusing first on the in-group of White defendants in the top row, the left-hand histogram shows sentencing dates to be spread evenly around 9/11, as expected (with the pregroup [postgroup] entirely to the left [right] of 9/11). The right-hand histogram shows the distribution of last offense dates. By design, both pre- and postdefendants committed their last offense before 9/11 and the distribution of last offense dates in the pre- and postgroup follow a similar shape, but the distribution for the postgroup is right shifted relative to the pregroup. The remaining panels in figure 3 show very similar patterns for sentencing and last offense dates for defendants in the two out-groups: Blacks and Hispanics.

The difference-in-differences empirical specification is given by

$$\begin{aligned} s_{iet} = \alpha &+ \sum_e \delta_e \mathit{Out\text{-}group}_e + \rho \mathit{Post}_t \\ &+ \sum_e \phi_e(\mathit{Out\text{-}group}_e \times \mathit{Post}_t) + \beta X_i + \gamma L_i \\ &+ \sum_f \omega_f \mathrm{OFF}_{if} + \sum_g \gamma_g G_{ig} + \sum_d \lambda_d D_{id} + \varepsilon_{iet}, \end{aligned} \tag{1}$$

where s_{iet} is the sentencing outcome for individual i of out-group e sentenced on day t based on a ±180-sentencing-day window around 9/11 and $Post_t$ is a dummy equal to 1 if the defendant comes up for sentencing post 9/11; all covariates (X_i, L_i, OFF_{if}, G_{ig}, D_{id}) are as described earlier. The term ε_{iet} is clustered by federal district. Our data do not contain judge identifiers, so we do not control for judge fixed effects.

D. Identifying Assumptions and Interpreting ϕ_e

Three assumptions underpin ϕ_e identifying a causal effect of 9/11 on sentencing outcomes for out-group e. First, the time defendants spend in the CJS between when they commit their last offense and when they come up for sentencing should not be differentially impacted by 9/11 across groups.

differences in victimization, the outcome variable is divided by its counterpart from he same month in the previous year and so can be interpreted as a growth rate. A color version of this figure is available online.

Table 2
Balance within Race

	White			Black				Hispanic			
	Control: Before 9/11	Treatment: After 9/11	*p*-Value 1: Difference	Control: Before 9/11	Treatment: After 9/11	*p*-Value 1: Difference	*p*-Value 2: DiD	Control: Before 9/11	Treatment: After 9/11	*p*-Value 1: Difference	*p*-Value 2: DiD
Sample size	6,137	6,857		5,162	5,714			7,749	8,609		
Number of dependents	1.09	1.11	.552	1.67	1.71	.396	.740	1.81	1.87	.094	.357
	(1.42)	(1.41)		(1.84)	(1.82)			(1.76)	(1.79)		
Age	38.0	38.4	.164	31.5	32.0	.039	.750	31.9	32.4	.001	.754
	(12.2)	(12.0)		(9.21)	(9.26)			(9.27)	(9.19)		
Marital status:											
Single	.335	.338	.741	.526	.541	.201	.483	.329	.327	.786	.670
Married or cohabiting	.431	.436	.710	.340	.328	.253	.331	.512	.509	.668	.584
Divorced, widowed, or separated	.213	.210	.688	.111	.112	.807	.649	.103	.099	.567	.932
Education:											
High school graduate or below	.659	.641	.105	.773	.778	.642	.160	.820	.814	.597	.488
Some college/college graduate	.331	.351	.087	.221	.216	.603	.125	.096	.092	.424	.072
Defense counsel:											
Privately retained	.167	.165	.833	.078	.082	.491	.529	.081	.077	.565	.900
Court appointed	.167	.174	.389	.165	.157	.383	.193	.265	.282	.092	.422
Federal public defender	.137	.132	.402	.156	.152	.678	.880	.276	.259	.223	.420
Other	.004	.005	.670	.005	.004	.597	.422	.002	.001	.154	.347

Note.—The sample refers to all cases for which sentencing occurs within a 6-month window of September 11, 2001. For those defendants sentenced after September 11, 2001 (treatment), the last offense was committed before September 11, 2001, and if sentenced before September 11, 2001 (control), the last offense was committed at least 180 days before September 11, 2001. Means and standard deviations (in parentheses for continuous covariates) are shown. The first p-values (p-value 1) are from tests of equality of the statistic within ethnic group across the two samples, based on an ordinary least squares (OLS) regression that allows standard errors to be clustered by district. The second p-values (p-value 2) are from tests of equality on the pre-post difference for the ethnic group in question relative to the White group. This is based on an unconditional difference-in-differences (DiD) specification, estimated by an OLS regression that allows standard errors to be clustered by district.

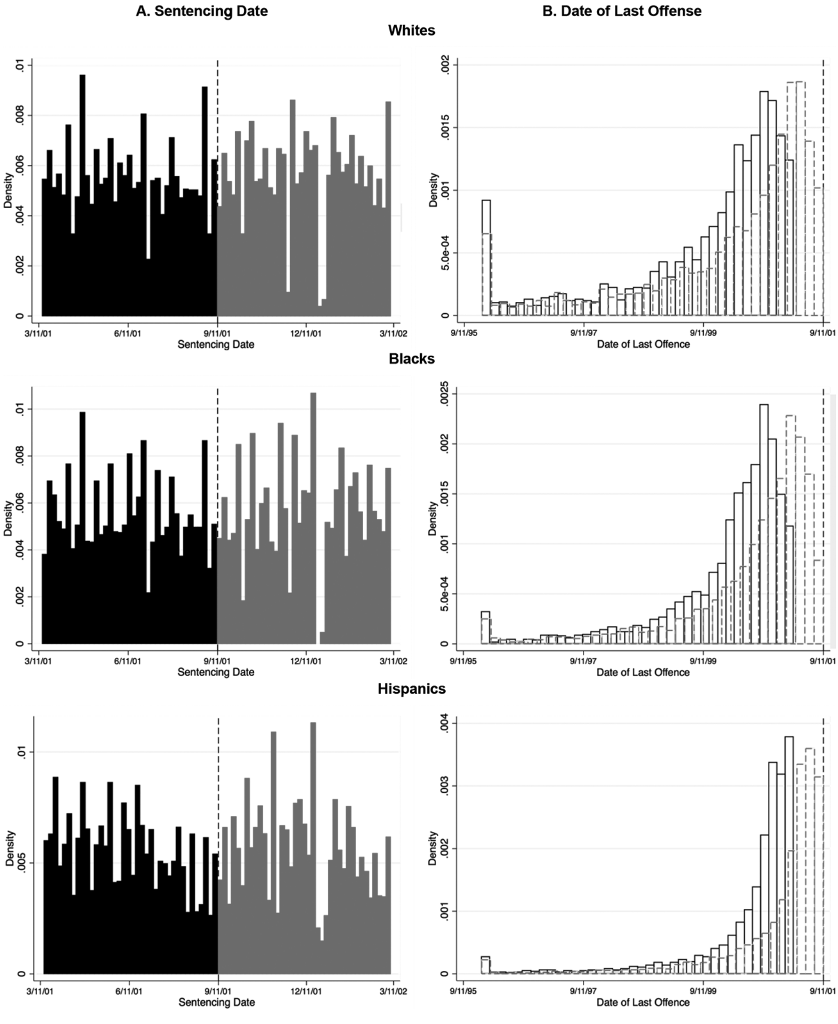

FIG. 3.—Sentencing and last offense dates. The left-hand graphs show the distribution of dates of sentencing date, for each group: 9/11 is indicated by the vertical dashed line. The right-hand graphs show the distribution of the dates of last offenses, by group. The first bar corresponds to a last offense date on or before January 1, 1996. The overlaid histograms are for those sentenced before and after 9/11. For those defendants sentenced after September 11, 2001, the last offense was committed before September 11, 2001, and if sentenced before September 11, 2001, the last offense was committed at least 180 days before September 11, 2001. A color version of this figure is available online.

This concern is ameliorated by there being proscribed periods of time between each stage of the federal CJS and restrictions on how long some stages can take (as shown in fig. 1). The evidence in figure 3 further points to there being no such queue jumping. We further address the concern using survival analysis to predict the time a defendant spends in the CJS between the date of last offense and sentencing by group. Second, we require there to be no race/ethnicity-time effects in sentencing differentials that naturally occur around 9/11 each year. We assess this via placebo checks using data from earlier years and also extend our preperiod to allow us to check for differential time trends across groups. Finally, we require there to be no missing covariates that determine sentencing outcomes, vary across groups, and change after September 11, 2001 (but not in placebo years). We address this issue by estimating bounds on the key difference-in-differences terms accounting for selection on unobservables.

Under these assumptions, ϕ_e still need not be interpretable as reflecting contagious/parochial animosity: it might reflect that judges anticipate changes in behavior of defendants after 9/11, with these expectations differing across out-groups. For example, 9/11 might have altered labor market outcomes for minorities, and this can affect recidivism rates differentially across groups (Orrenius and Zavodny 2009); alternatively, judges might anticipate that after 9/11 the police will reallocate resources in a way that differentially changes future detection probabilities by race/ethnicity. Taken together, such channels represent different forms of statistical discrimination, where stereotyping of defendants by group can lead to differential outcomes by race/ethnicity after 9/11.[10]

We use two strategies to narrow the interpretation: (i) decomposition analysis to show how much of the differential is attributable to changing sentencing penalties on observables and (ii) correlating sentencing differentials to judge characteristics, including race/ethnicity, in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using police arrest data (Anwar and Fang 2006; Park 2017).

IV. Judges and Sentencing Outcomes

A. Downward Departures

Table 3 presents estimates of equation (1) for downward departures, the key margin of judicial discretion at sentencing. Column 1 shows that Hispanic-White sentencing gaps open up after 9/11: relative to Whites, the likelihood that Hispanics receive a downward departure falls significantly by 3.8 percentage points. We see no such impact on Black defendants, on whom the post-9/11 impact for downward departures is a precisely estimated zero (and as shown in the column, this is significantly different from the post-9/11 impact on Hispanics; $p = .041$). Recall that as shown in table 1, no Hispanic-White differential in rates of downward departure existed before 9/11. This Hispanic-White sentencing differential opens up only after 9/11. If 9/11

[10] Of course, statistical discrimination is not legally permissible because sentencing differentials cannot be justified on the basis of statistical generalizations about group traits, irrespective of whether there is an empirical foundation for this (*JEB v. Alabama ex rel TB*, 511 US 127 1994).

Table 3
Judges' Downward Departure Decisions around 9/11 (Dependent Variable: Downward Departure Granted by Federal Judge)

	Baseline (1)	Reason: Criminal History Category Overrepresented (2)	Reason: Pursuant to Plea Bargain (3)	Reason: General Mitigating Circumstances (4)	Reason: Other (5)	Initial Arrest Codes (6)
Sentenced pre-9/11 × Hispanic	−.038***	−.013***	−.011	−.001	−.013**	−.046***
	(.010)	(.003)	(.007)	(.007)	(.007)	(.016)
Sentenced pre-9/11 × Black	−.013	−.005	.002	−.003	−.007	−.013
	(.008)	(.004)	(.003)	(.003)	(.005)	(.011)
Sentenced pre-9/11	.006	.003	−.000	.001	.002	.003
	(.007)	(.002)	(.002)	(.004)	(.004)	(.009)
Offender, legal, and district controls	Yes	Yes	Yes	Yes	Yes	Yes
Offense type codes	Final	Final	Final	Final	Final	Arrest
Guideline cells	Yes	Yes	Yes	Yes	Yes	No
p-value (post-9/11 × Black = post-9/11 × Hispanic)	.041	.036	.123	.757	.351	.079
Implied sentence length impact (Hispanic)	.736					.889
% of pre-9/11 sentence differential	18					29.8
Adjusted R^2	.256	.042	.289	.068	.135	.257
Observations	40,228	40,228	40,228	40,228	40,228	26,852

NOTE.—Ordinary least squares regression estimates are shown in all columns. Standard errors clustered by district are reported in parentheses. In cols. 1–5, the sample of 40,228 federal cases is used (those that come up for sentencing in a 6-month window on either side of September 11, 2001). For those defendants sentenced after September 11, 2001, the last offense was committed before September 11, 2001, and if sentenced before September 11, 2001, the last offense was committed at least 180 days before September 11, 2001. Columns 2–5 code downward departures into various broad categories of how judges justify their decision to depart. In col. 6 the sample is restricted to those cases that can be linked back to arrest (stage 0). The dependent variable throughout is a dummy for whether the case receives a downward departure (where in cols. 2–5 this is modified on the basis of the reasons given for departure). In all columns we condition on defendant group (White, Black, Hispanic), whether the case comes up after 9/11, and interactions between the two as well as the following: offender characteristics (dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this nonmissing-age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a nonmissing-dependents dummy), legal controls (a dummy for whether information on the defense counsel is missing and a nonmissing dummy interacted with the type of defense counsel [privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements]), the guideline cell, and federal district dummies. In cols. 1–5 we control for the primary offense type and a dummy for multiple offenses. In col. 6 we instead control for arrest offense codes but not guideline cells. The p-value in each column is for the null that the coefficients on the post-9/11 × Black and post-9/11 × Hispanic dummy interactions are equal against a two-sided alternative.
** Significant at 5%.
*** Significant at 1%.

sparked a rise in animosity toward Muslims, this pattern of results across out-groups is consistent with judges displaying contagious animosity from Muslims to Hispanics, while their social preferences are independent between Muslims and Blacks.

Judges have to provide an explanation for downward departures: columns 2–5 code these into broad categories. The differential impact on Hispanics is driven by judges being less likely to downward depart as a result of (i) a belief that the criminal history of the defendant is overrepresented or (ii) other reasons. For the first type of downward departure, the post-9/11 impact on Hispanics is significantly different from that on Blacks ($p = .036$). There is no statistically significant shift in downward departures related to general mitigating circumstances and no precisely estimated impact on downward departures related to plea bargains.

A greater Hispanic-White sentencing differential after 9/11 could be the result of either contagious animosity, where anti-Muslim sentiment hurts Hispanics, or parochial animosity, where anti-Muslim sentiment increases in-group altruism toward Whites. The evidence rules out the latter interpretation because (i) the post-9/11 indicator on the likelihood of downward departure (for Whites) is a precisely estimated zero and (ii) we find statistically significant differences in the impacts between Hispanic and Black defendants, again suggesting that the results are not driven by increased altruism toward the White in-group.

We can convert the impacts on the propensity to downward depart into an implied change in expected sentence length as follows. To do so, we calibrate sentence length impacts assuming that the only channel through which 9/11 impacts sentence length is the likelihood of downward departure and so hold constant other channels, such as (i) the number of guideline cells shifted conditional on downward departure and (ii) sentence length within guideline cell conditional on no departure. We return to these other channels below.

For the current exercise, we denote the probability of being assigned to guideline cell g as p_g, the probability of being downward departed as p_d, and the expected sentence conditional on being sentenced within the range of guideline cell g as $E[s|g]$. The implied change in expected sentence length is

$$\textstyle\sum_g p_g \Delta p_d \{E[s|g-4] - E[s|g]\}, \tag{2}$$

where we (i) use the pre-9/11 empirical distribution of defendants (in a given out-group) across guideline cells to measure p_g, (ii) assume that an individual moves four guideline cells (to $g - 4$) if downward departed (which is true for the median defendant before 9/11), and (iii) take the cell g midpoint to estimate $E[s|g]$. Column 1 in table 3 shows the implied impact on Hispanic sentence lengths to be .736 months, corresponding to 18% of the conditional pre-9/11 Hispanic-White differential in sentence length (table 1, col. 4).[11]

[11] The formula for the implied sentence length impact is justified given that the downward departure impact on Hispanics occurs across regions of the guideline

To monetize these sentencing impacts, we note that (i) the marginal annual cost per year of imprisoning a male prisoner is $29,000 (CRS 2013) and (ii) in the federal system, the elasticity of incarceration with respect to sentence is approximately 0.87 (Rehavi and Starr 2014). Combining these with our implied sentence impact suggests that 9/11 lead to an increase of $1,547 in incarceration costs per Hispanic defendant, mapping to a large increase in total costs of the federal CJS given that 40% of all defendants are Hispanic.[12]

The analysis conditions on the offense type the defendant is charged with. This replicates earlier work in economics on sentencing outcomes, by conditioning on all information available to judges at the point they make their key decision. An alternative approach, following Rehavi and Starr (2014), is to condition only on observables determined at the point a defendant enters the federal CJS. The justification for doing so is that prosecutors might manipulate the offense level, say through selective fact-finding and perhaps in anticipation of a judge's behavior (Schanzenbach and Tiller 2007; Cohen and Yang 2019). To address this issue, we exploit information from the arrest stage of the criminal timeline (stage 0): for the 67% of cases that can be linked back to the arrest stage, we condition on more than 400 codes corresponding to the precise offense the defendant was originally arrested for (rather than conditioning on the 31 offense type codes or 258 guideline cells based on prosecutor decisions during the timeline). Column 6 shows that conditional on original arrest codes, the Hispanic-White differential after 9/11 on downward departures remains significant and is larger in absolute value at −.046 percentage points. This impact remains statistically different from any post-9/11 impact on Black defendants ($p = .079$), and the implied sentence length impact is 0.889 months, nearly 30% of the conditional pre-9/11 Hispanic-White sentence differential.

B. Sentence Length

We next consider sentence length as the outcome s_{iet}. The calibration exercise in equation (2) assumed that the only channel through which 9/11 impacts sentence length is the likelihood of downward departure, holding constant other channels, such as (i) the number of guideline cells shifted conditional

cell table in fig. A1. The impact for Hispanic defendants assigned to region A (so with relatively low offense severity and criminal history scores) is −0.036, while for Hispanic defendants in regions B–D the impact is −0.037, with both estimates being statistically significantly different from zero and significantly different from the post-9/11 impacts on Blacks ($p = .041$ and .058 respectively).

12 Mueller-Smith (2016) estimates the total social cost generated by 1 year of incarceration to be between $56,000 and $66,000. An alternative benchmark is how sentencing differentials in the federal CJS have been impacted by institutional reforms. For example, sentencing guidelines were abolished in 2005 following the Supreme Court's decision in *United States v. Booker*. There is mixed evidence on what impact this abolition had on sentencing differentials. Fischman and Schanzenbach (2012) report no effects, while Yang (2015) uses individual matched judge and defendant data and finds that sentences for Blacks rise by 2 months as a result. Hence, the magnitude of our main effect arising from contagious animosity corresponds to just over one-third of this.

on downward departure (which from table 1 we see applies to 17% of defendants) and (ii) sentence length within guideline cell conditional on no departure (which applies to the remaining 83% of defendants). Measuring an overall Hispanic-White sentence differential is complicated by the fact that a small share of defendants are impacted through downward departures, and channels i and ii above might move in opposite directions.

Notwithstanding this issue, to begin with table 4 shows impacts on overall sentence length (in months) from estimating equation (1). Column 1 shows $\hat{\phi}_H$ not to be statistically different from zero. In column 2 we remove defendants with a life sentence (as these are all top coded at $s_{iet} = 470$ months). The point estimate of $\hat{\phi}_H$ then becomes positive but is still not different from zero. To make the results less sensitive to impacts on the tails of the distribution of sentence lengths caused through channels i and ii above, column 3 shows estimates from a quantile regression at the median sentence length, following the approach of Firpo, Fortin, and Lemieux (2007). The point estimate of the Hispanic-White sentencing differential rises to $\hat{\phi}_H = 715$, closely matching the calibrated

Table 4
Judges' Sentencing Decisions around 9/11 (Dependent Variable: Sentence Length [Months])

	Full Sample	Removing Life Sentences	
	OLS (1)	OLS (2)	Quantile (Q50) (3)
Sentenced post-9/11 × Hispanic	−.367	.056	.715
	(.712)	(.595)	(.629)
Sentenced post-9/11 × Black	.400	.027	−.396
	(.938)	(.783)	(.560)
Sentenced post-9/11	.873**	.762*	.368
	(.418)	(.407)	(.446)
Offender, legal, and district controls	Yes	Yes	Yes
Offense type codes	Final	Final	Final
Guideline cells	Yes	Yes	Yes
p-value: post-9/11 × Black = post-9/11 × Hispanic	.432	.971	.062
Adjusted R^2	.754	.773	.720
Observations	40,228	40,116	40,116

NOTE.—Standard errors clustered by district are reported in parentheses. In col. 1 the full sample of 40,228 federal cases is used (those that come up for sentencing in a 6-month window either side of September 11, 2001). For those defendants sentenced after September 11, 2001, the last offense was committed before September 11, 2001, and if sentenced before September, 11, 2001, the last offense was committed at least 180 days before September 11, 2001. Columns 2 and 3 drop life sentences (which are top coded at 470 months). Column 3 presents quantile regression estimates at the median. In all columns we condition on defendant group (White, Black, Hispanic), whether the case comes up after 9/11, and interactions between the two as well as the following: offender characteristics (dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this nonmissing-age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a nonmissing-dependents dummy), legal controls (a dummy for whether information on the defense counsel is missing and a nonmissing dummy interacted with the type of defense counsel [privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements]), the primary offense type and a dummy for multiple offenses, the guideline cell, and federal district dummies. The p-value in each column is for the null that the coefficients on the post 9/11 × Black and post 9/11 × Hispanic dummy interactions are equal against a two-sided alternative.
* Significant at 10%.
** Significant at 5%.

sentence length of 0.736 (which assumed no impacts within cells or in cell movements). We reject the null that the differential effects of 9/11 on sentence lengths for Hispanics and Blacks are the same ($p = .062$).

To build a more complete picture of the sentence impacts of 9/11 that also sheds light on channels i and ii, we next define a sentence adjustment for defendant i initially assigned to guideline cell G: $sa_{iG} = s_i - \min(s_{iG})$. Negative values of sa_{iG} represent a final sentence below the guideline cell range (which arises from a downward departure), $sa_{iG} = 0$ represents the sentence being at the lower bound of the guideline cell (which is a natural focal point for sentence length, with 33% of sentences being at this bound before 9/11), and positive values represent a higher sentence within the guideline cell (which could also be due to a binding statutory minimum sentence length requirement). We then estimate specifications analogous to equation (1), where the outcome variable is $\Pr(sa_{iG} \leq \tau)$ where $\tau = -1, -2, \ldots, -24$, $\Pr(sa_{iG} = 0)$, and $\Pr(sa_{iG} \geq \tau)$ where $\tau = 1, 2, \ldots, 12$. The asymmetry reflects that downward sentence adjustments of up to 2 years are far more common than upward sentence adjustments beyond 12 months of the guideline cell minimum. We note that excluding life sentences, the average width of a guideline cell is 15 months.

The resulting sequence of difference-in-differences estimates is shown in figure 4. The top panels show the estimated Hispanic-White differential for each sentence adjustment and the corresponding 95% confidence intervals. The left-hand panel does so unconditionally; the right-hand panel controls for the full set of covariates in equation (1).

For sentence adjustments below the minimum of the guideline cell ($\tau < 0$) we see that (i) Hispanic defendants are significantly less likely to have sentence adjustments between -9 and -1 months (the range is slightly larger when we do not conditional on covariates) and (ii) there is no significant differential impact of 9/11 on sentence adjustments below this level ($\tau \leq -10$). This result suggests that the marginal Hispanic defendant less likely to be downward departed after 9/11 is in a sentencing adjustment band just below the minimum of their original guideline cell. Defendants farther away from this minimum to begin with are inframarginal and are not differentially impacted by 9/11.

The right-hand side of each panel provides an indication of where the marginal Hispanic defendant is then shifted to: for sentence adjustment at or above the minimum of the guideline cell ($\tau \geq 0$) we see an increased mass of defendants precisely at the minimum of the guideline cell ($\tau = 0$), with declining impacts for conditional sentence adjustments of 1 month and above.

The lower panels of figure 4 repeat the analysis for Black-White sentencing adjustment differentials. Both the unconditional estimate and the conditional estimate are smaller in magnitude and not ever statistically different from zero.

As a final step of analysis, we focus in on the resulting impacts on sentence lengths from these changes in sentence adjustments. Our approach is to try to identify those defendants who in the counterfactual absent 9/11 would have been most likely to be downward departed and then measure their sentence

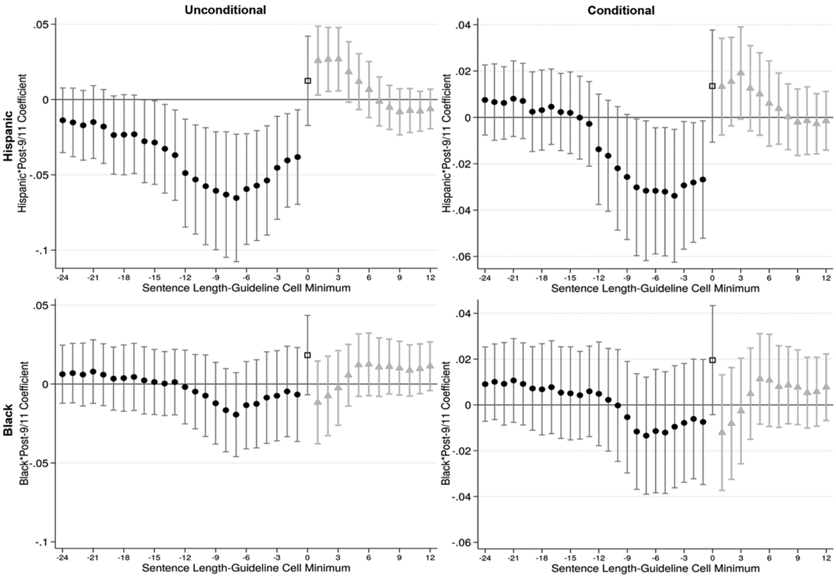

FIG. 4.—Sentencing adjustments. The graphs show estimates from different regressions, where each coefficient and corresponding 95% confidence interval comes from a separate regression. The sample of 40,228 federal cases is used (those that come up for sentencing in a 6-month window either side of September 11, 2001). For those defendants sentenced after September 11, 2001, the last offense was committed before September 11, 2001, and if sentenced before September 11, 2001, the last offense was committed at least 180 days before September 11, 2001. Standard errors are clustered by district. The regressions are based on the difference between an individual's sentence length and the minimum sentence length recommendation in their allocated guideline cell. On the basis of this difference, we create a series of dummy variables, which are the dependent variables in the graphs. The first set take a value of 1 (0 otherwise) if the difference in sentence length–guideline cell minimum is less than or equal to a negative integer in the range −24 to −1 (the estimates based on these dependent variables are represented by solid black circles). We treat zero separately, creating a dummy if sentence length equals the guideline cell minimum (corresponding estimates for this dependent variable are represented by hollow black squares). Finally, we create a set of dummy variables that take a value of 1 (0 otherwise) if the sentence length–guideline cell minimum is greater than or equal to a positive integer in the range 1 to 12 (estimates for which are represented by solid triangles). We then run a separate ordinary least squares regression based on each of these dependent variables and estimate difference-in-differences models, both without and without a set of additional control variables. In the unconditional models we condition on defendant group (White, Black, Hispanic), whether the case comes up after 9/11, and interactions between the two. In the conditional models we include the regular set of controls. Estimates for the two difference-in-differences terms post-9/11 × Hispanic and post-9/11 × Black are presented. A color version of this figure is available online.

differential after 9/11 against this counterfactual. We proceed as follows. First, we use the entire pre-9/11 sample (back to October 1998) to estimate the likelihood of a downward departure using the same covariates as in equation (1) but allowing for more detailed categorizations of age and the number of dependents (because the sentencing guidelines make explicit that downward departures can occur partly based on family responsibilities or prior good works). We estimate this prediction model using a probit specification and do so separately by outgroup *e*. We then take our baseline working sample of defendants up for sentencing in the window around 9/11 and group defendants into percentile bands of their predicted probability of downward departure, $\hat{p}_{\mathrm{DD}}$, based on the pre-9/11 models. In each subsample, we keep observations if the predicted probability exceeds any given percentile value, so moving from the 5th to the 90th percentile we progressively keep fewer observations. Based on each of these subsamples, we run our standard difference-in-differences specification where the dependent variable is sentence length. Finally, we plot the difference-in-differences for these percentile subsamples of $\hat{p}$ along with their corresponding 95% confidence intervals and overlaid with the histogram of $\hat{p}$.

The results are shown in figure 5*A*. We see that for defendants between the 70th and 85th percentiles of the predicted probability of downward departure, there is a significant increase in sentence lengths. The magnitude of this effect is just over 2 months. Consistent with the results on sentence adjustments, we see that defendants with the highest predicted probability of being downward (more than the 90th percentile of $\hat{p}$) have no change in the sentence outcome—as figure 4 showed, they are not the marginal defendant differentially impacted by 9/11. Second, we see that the majority of defendants—those below the 70th percentile or above the 90th percentile of $\hat{p}$—have no significant impact on their sentence length, and this is line with 83% of them not being subject to downward departures (table 1). This is what mutes the overall impact on sentence lengths shown in table 4.

Figure 5*B* shows the findings if the first-stage prediction model for the likelihood to be downward departed includes additional interactions between the number of children and the six broad categories of criminal history shown in table A1.

How large is a 2-month impact on sentence length? It corresponds to 50% of the conditional Hispanic-White sentencing gap before 9/11 shown in table 1. It is also comparable in magnitude to the sentencing impacts documented in Yang (2015), who studied racial sentencing differentials once sentencing guidelines were struck down in 2005. She finds that increasing judicial discretion in sentence lengths increased average sentence lengths for Black defendants relative to Whites by 2 months. Hence, our findings suggest that the impact on sentence lengths arising through social preference structures and contagious animosity around 9/11 being transmitted from Muslims to

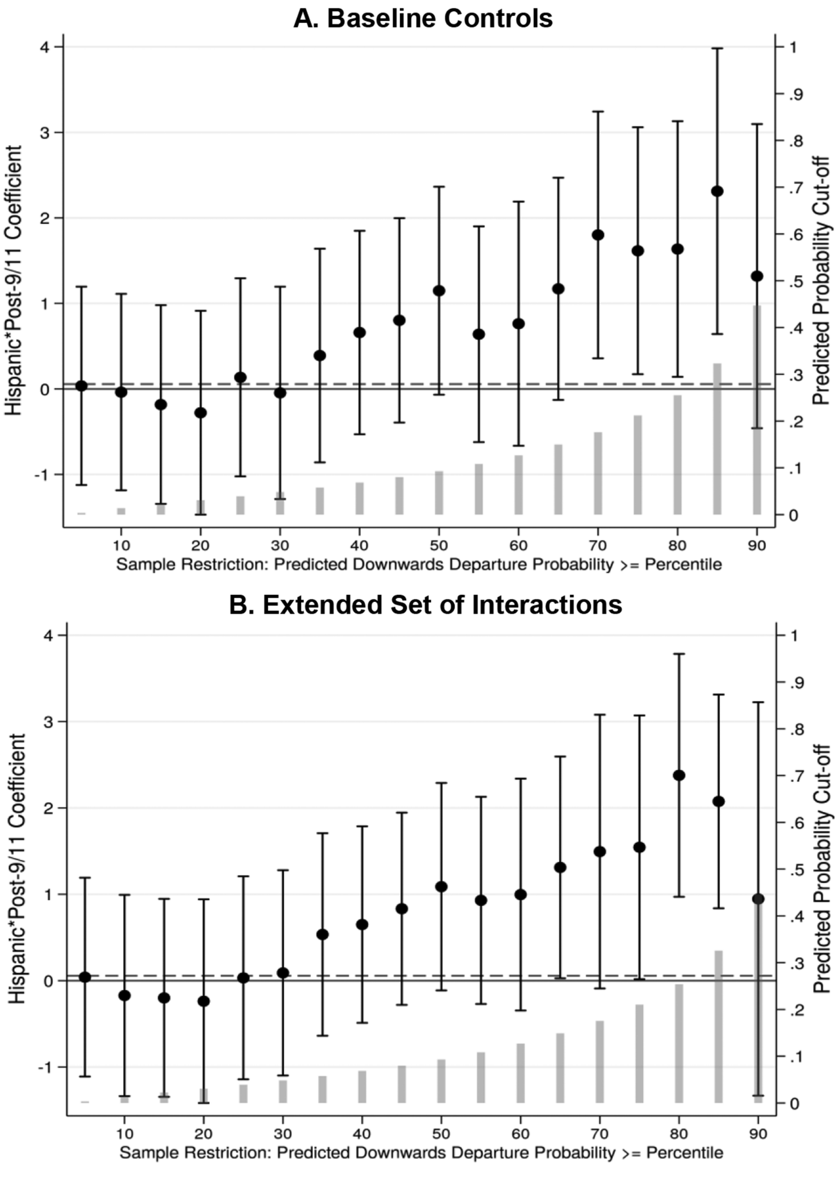

FIG. 5.—Predicted impact on sentence length (months). Each panel shows estimates where each coefficient and corresponding confidence interval comes from a separate difference-in-differences regression. The regressions are based on different subsamples of the baseline sample of 40,116 federal cases. The subsamples are created as follows. We begin with an expanded sample of all non-life-sentence federal cases that come up for sentencing between October 1, 1998, and 180 days after September 11, 2001. For cases sentenced before 9/11, we run a probit regression by

Hispanics is around the same magnitude as that arising from an institutional change in sentencing policy on Black defendants.

C. Citizenship and Offense Type

There are two obvious reasons why Hispanic-White sentencing differentials might become exacerbated after 9/11 while Black-White differentials remain unchanged that have nothing to do with contagious animosity across out-groups. The first is that Hispanics constitute the majority of non-US-citizen defendants. Punishments for noncitizens, such as deportation, differ from those available for citizens and residents/legal aliens, and these might become harsher for noncitizens after 9/11. If so, the Hispanic-White differential would just pick up this differential selection into citizenship status.

Seventy-one percent of defendants are citizens, 43% of Hispanic defendants are citizens, and 91% of noncitizens are Hispanic. Given this close alignment between race and citizenship status, it is hard to cleanly separate the two, but we do so to the extent the data allow. Column 1 of table 5 allows impacts to vary between Hispanics citizens (US citizen, resident/legal alien) and Hispanic noncitizens (illegal alien, non-US citizen, status unknown). For both groups of Hispanic, those who are sentenced after 9/11 are significantly less likely to be downward departed. For Hispanic citizens, the impact is a 2.8 percentage point reduction in the likelihood of a downward departure, corresponding to an implied higher sentence length of 17% of the pre-9/11 Hispanic citizen–White differential. For Hispanic noncitizens, the impact is a 4.4 percentage point reduction in downward departure, an implied sentence length increase mapping to 16% of the pre-9/11 Hispanic noncitizen–White

---

ethnicity where the dependent variable is a dummy for downward departure. *A* is based on our regular set of controls. *B* is based on the same set of controls but additionally controls for a set of dummies based on an interaction between number of dependants and criminal history category. We use this first-stage regression to predict the probability of a downward departure for the full, expanded sample (i.e., including post-9/11) and then restrict the sample to the 180-day window around September 11, 2001. We use this predicted probability to create the subsamples on which the sentence length regressions are based. We calculate the percentiles of the predicted probability of downward departure for values from 5 to 90 in increments of 5. We subsequently keep observations if the predicted probability exceeds this percentile value. Based on each subsamples, we run a difference-in-differences regression where the dependent variable is sentence length and the regular set of control variables are included. Point estimates and corresponding 95% confidence intervals are shown for the post-9/11 $\times$ Hispanic term. On the right-hand *Y*-axis we show the value of the predicted probability at each percentile cutoff. In each graph, the dashed line represents the difference-in-differences estimate based on our working sample around the 9/11 window, excluding defendants with life sentences. A color version of this figure is available online.

sentencing differential. There is no statistical difference between the two impacts ($p = .278$).

A second reason why Hispanic-White sentencing differentials might increase after 9/11 is that they are more likely to be charged with immigration offenses. If such offenses are more severely punished after 9/11, ϕ_H might just pick up that Hispanics are charged with immigration offenses at a greater rate than others. To address the issue, the remaining columns of table 5 split the sample by offense type (drug, immigration, other) while still allowing the impact of ethnicity to vary between Hispanic citizens and Hispanic noncitizens. For immigration offenses, the vast majority of defendants in the federal system are Hispanic (either citizens or noncitizens). Hence, when examining those offenses we restrict the sample further to Hispanics only.

Across offense types, we find no significant differences between the impact of 9/11 on Hispanic citizens and that on noncitizens: (i) Hispanic noncitizens are significantly less likely to receive downward departures for drug offenses (col. 2), but this effect is not different from that for Hispanics citizens ($p = .210$); (ii) on immigration offenses, there is little robust evidence that Hispanics, either citizens or noncitizens, experience a change in the likelihood of receiving a judicial downward departure, and this remains the case if we focus exclusively on border states (cols. 3, 4); and (iii) the lower likelihood of downward departures after 9/11 is largely driven by the impact on Hispanic citizens for other offenses (col. 5), but again this is not different from that for Hispanics citizens ($p = .722$). These constitute around 40% of all offenses and often relate to firearms.[13]

Table A3 shows that these results by offense type continue to hold when we use the original arrest codes from the start of the criminal timeline (stage 0): we find no robust evidence that sentencing differentials for drug, immigration, or other offenses change differentially after 9/11 between Hispanic citizen and Hispanic noncitizen defendants.

D. Robustness and Support for Identifying Assumptions

Tables A4–A6 conduct a battery of robustness checks on our core finding from table 3. These show the result to be robust to (i) alternative levels of clustering standard errors; (ii) exclusion of cases where perhaps because of a prosecutor's decision-making over the initial offense charges filed (stage 3 in fig. 1), statutory minima, or maxima bind partially over the range set by the guideline cell (Rehavi and Starr 2014); and (iii) estimating equation (1) separately for each group. We also combine information on Hispanic origins and race to examine whether our findings pick up ethnic, rather than racial, sentencing differentials. In each set of robustness checks, we find that the

[13] In line with our results, Mustard (2001), using data on federal criminal cases, documents that the Hispanic-White sentence gap is generated by those convicted of drug trafficking and firearm possession/trafficking.

results hold irrespective of whether we control for final offense codes or initial arrest codes.

The appendix also provides evidence in support of the three identifying assumptions required to interpret ϕ_e as measuring a causal impact: (i) table A7 shows the main results to be robust to controlling for time of offense (irrespective of whether we use final offense codes or initial arrest codes) and (ii) table A8 uses survival analysis to show that the time defendants spends between their last offense and when they come up for sentencing is not differentially impacted by 9/11 across groups.

We next address the concern that there are race/ethnicity-time effects in sentencing differentials that naturally occur around 9/11 each year. We do so using four pieces of evidence. First, we use data from earlier years to construct placebo 9/11 effects. As table A9 shows, the impact for Hispanics on downward departures occurs after 9/11 in 2001. Again, this result is robust to controlling for either final offense codes or initial arrest codes. Second, we check for pretrends by considering all offenses committed before 9/11 (even if the defendant has been sentenced before 9/11 and exited the system). We thus define the preperiod as starting from October 1998. In this extended sample, we can control for linear time trends in rates of downward departure, which can vary by group. Table A10 shows that our core result remains robust: there remains a significant fall in the likelihood of Hispanic defendants being downward departed after 9/11 (col. 3). The magnitude of the effect is −0.042 (standard error = 0.012), which is near identical to be baseline estimate of −.038 (standard error = 0.010). This is over and above the long-run upward trend in the likelihood of Hispanics being downward departed shown (and the magnitude of this trend is slight [.002]).

Third, we address concerns that impacts are driven by the Patriot Act, which was enacted 45 days after 9/11. To shed light on the matter, we estimate a dynamic specific analogous to equation (1) that estimates impacts in 15-day windows after 9/11. As we showed earlier, immigration offenses do not drive the main result; figure A2 documents how impacts on judicial departures for Hispanics appear after 9/11 and before and after the Patriot Act for offenses unrelated to the Patriot Act. We find that the point estimates are of similar magnitude to the main estimate from equation (1) and are relatively stable over each of these 15-day windows, including those before the Patriot Act was introduced.

Fourth, we collect data on the date of confirmation of Bush-appointed US attorneys (shown in fig. A3), to establish in table A11 that none of the post-9/11 impacts we measure are driven by the share of time a federal district spends under a Bush-appointed US attorney, which might otherwise signal a change in how the CJS views the trade-off between justice and social protection. Again, this is robust to controlling for either final offense codes or initial arrest codes.

Table 5
Citizenship and Offense Type (Dependent Variable: Downward Departure Granted by Federal Judge)

	All Offenses (1)	Drug Offenses (2)	Immigration Offenses: Hispanics Only (3)	Immigration Offenses: Hispanics Only, Border States (4)	All Other Offenses (5)
Sentenced post-9/11 × Hispanic citizen	−.028**	−.017	−.054	−.038	−.031**
	(.011)	(.013)	(.037)	(.049)	(.014)
Sentenced post-9/11 × Hispanic noncitizen	−.044***	−.054*	.033	.017	−.018
	(.013)	(.028)	(.037)	(.048)	(.032)
Sentenced post-9/11 × Black	−.013	−.003			−.018*
	(.008)	(.014)			(.010)
Sentenced post-9/11	.005	−.001			.009
	(.007)	(.013)			(.008)
Offender, legal, and district controls	Yes	Yes	Yes	Yes	Yes
Offense type codes	Final	Final	Final	Final	Final
Guideline cells	Yes	Yes	Yes	Yes	Yes
Implied sentence length impact (Hispanic, citizen)	.575	.520	.741	.478	.367
	[17.2%]	[9.2%]			[19.0%]
Implied Sentence Length Impact (Hispanic, noncitizen)	.821	1.372	.424	.422	.151
	[15.9%]	[18.1%]	[18.1%]	[15.5%]	[5.0%]
p-value: post-9/11 × Hispanic citizen = post-9/11 × Hispanic noncitizen	.278	.210	.237	.583	.722
Adjusted R^2	.258	.292	.357	.342	.091
Observations	39,937	18,222	6,147	4,534	14,978

Note.—Ordinary least squares regression estimates are shown throughout. Standard errors clustered by district are reported in parentheses. The sample of 39,937 federal cases is used (those that come up for sentencing in a 6-month window either side of September 11, 2001) and for which defendant citizenship is not missing. For those defendants sentenced after September 11, 2001, the last offense was committed before September 11, 2001, and if sentenced before September 11, 2001, the last offense was committed at least 180 days before September 11, 2001. Column 1 covers all offenses. Columns 2–5 are restricted to drug, immigration, and other offenses, respectively, where for immigration offenses only Hispanic defendants are included and col. 4 further restricts the sample to US-Mexico border states. The dependent variable is a dummy for whether the case receives a downward departure. In all columns we condition on interactions between Hispanic ethnicity, defendant citizenship (where citizens are defined as being US citizens or resident/legal aliens and noncitizens are illegal aliens, non-US citizens, and those for whom alien status is unknown), and whether the case comes up after 9/11 as well as each of these control variables alone. In all specifications the regular set of controls are included. The percentage reported in brackets is the percentage of the pre-9/11 differential the implied sentence length impact corresponds to. The p-value in each column is for the null that the coefficients on the post-9/11 × Hispanic citizen and post-9/11 × Hispanic noncitizen dummy interactions are equal against a two-sided alternative.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

The final identifying assumption required is that there are no missing covariates that determine sentencing outcomes, vary across groups, and change after September 11, 2001 (but not in placebo years). We address this following Altonji, Elder, and Taber (2005) and Oster (2019) to estimate bounds on the treatment effect of $Out\text{-}group_e$, accounting for selection on unobservables. The results in table A12 show that these bounds on $\hat{\phi}_e$ are tight. For them to include zero requires unobserved factors changing for Hispanics after 9/11 that are orders of magnitude more predictive of sentencing outcomes than the covariates in equation (1), including the full set of guideline cell dummies.

V. Prosecutors and Presentencing Outcomes

Prosecutors represent a second crucial actor determining defendant outcomes. We extend our research design to examine the presentence prosecutorial decision-making. This enables us to provide insight into whether prosecutors, who around 9/11 were overwhelmingly White, display behaviors toward out-groups consistent with the results found for judges.[14]

Prosecutors decide the initial offense charge filed against defendants (stage 3 in fig. 1). In the federal criminal code, definitions of crimes often overlap, providing prosecutors discretion over initial charges. These charges are crucial because they determine (i) whether statutory minima/maxima sentences bind and take precedence over guideline cell sentence ranges and (ii) outside options in plea bargaining (defendants might plead to a lesser charge to avoid being charged with an offense with a mandatory minimum; Yang 2016).[15]

In table 6, we use the pre-9/11 sample to first document, by out-group, (i) the frequency with which defendants receive an initial charge with a nonzero statutory minimum sentence and (ii) the length of statutory minimum sentence associated with their initial offense (setting initial offense charges without a statutory minimum to zero).[16] Before 9/11, (i) Blacks are unconditionally 23.3 percentage points more likely to be charged with an offense with a

[14] A recent study of state prosecutors by the Women Donors Network found that (i) 95% of elected prosecutors are White and (ii) the majority of states have no elected Black prosecutors. A summary of the findings are available at https://wholeads.us/wp-content/uploads/2019/03/Justice-For-All-Report_31319.pdf.

[15] Many forms of statutory minima exist and can have precedence over the minimum from the guideline cell. In 15.8% (3.6%) of cases, the statutory minimum is above (below) the guideline minimum (maximum).

[16] Our coding of statutory minimums differs from the primary coding in Rehavi and Starr (2014). They derive minima on the basis of initial offense charges, while we use the realized mandatory minima as recorded from the MFCS data. To gauge the relationship between the two codings, we use the Administrative Office of the US Courts stage of the FJSRC data to create a marker for whether there is a change in offense between the initial charge and the conviction state using three increasingly detailed descriptions of offense: (i) most serious offense category (of which there are 51 distinct values), (ii) most serious offense (204 distinct values), and (iii) primary

statutory minimum sentence length (col. 1) and (ii) conditional on offender and legal counsel characteristics and federal district, Blacks and Hispanics are significantly more likely to be charged with offenses with a statutory minimum (col. 2). We next condition on a rich set of codes corresponding to the original offense the defendant was arrested for. The result in column 3 shows that when doing so, there remain significant Black-White and Hispanic-White differences in the likelihood of a nonzero statutory minimum offense charge being given.

Columns 4–6 document that these differences translate into a similar pattern of differentials before 9/11 for statutory minimum sentence lengths. Blacks receive charges carrying minimum sentences that are conditionally 22 months longer than those for Whites, falling to 7.8 months in cases linked to arrest offense codes. For Hispanics, prosecutors set initial charges with associated statutory minimums that are 14 months longer (or 63% higher) than those for Whites, falling to 7.4 months in cases that can be linked to arrest offense codes.

We next use our research design to examine whether 9/11, which increased animosity toward Muslims, had spillover effects on other out-groups in the federal CJS through prosecutors' decisions. We consider a narrow window covering a cohort of 3,600 defendants, all of whom entered the federal system before 9/11 but had their initial offense charges filed either side of 9/11. Taking the date of last offense to proxy for time of entry into the federal CJS (stage 1), we exploit the fact that the system requires defendants in (out of) custody to have their initial offense charges brought within 14 (21) days. This allows us to define two groups of defendant: (i) those whose last offense was committed 29–42 (43–63) days before 9/11 (depending on whether they are in custody) and so whose initial offense charge was determined before 9/11 and (ii) those whose last offense was committed 14 (21) days before 9/11 until the day before 9/11 and so their initial offense charge would have been determined just after 9/11. We estimate a specification analogous to equation (1) but where the outcomes are (i) whether the defendant receives an initial charge with a nonzero statutory minimum sentence and (ii) the length of statutory minimum sentence associated with their initial offense. We do not condition on final offense type or the later determined guideline cell.[17]

offense charge (1,543 distinct values). Of the defendant sample we can match from sentencing back to the arrest data, the coding of offenses was unchanged for 93.4% of cases under definition i, 88.6% under definition ii, and 81.6% under definition iii.

[17] We remove those whose last offense was committed 15–28 (22–42) days before 9/11 to avoid misclassifying individuals. If we condition on arrest offense codes followed by the combination of a smaller sample and a rich set of arrest codes to control for means, we lose precision, although the signs of all *Post* × *Hispanic* interactions remain the same as those shown.

Table 6
Pre-9/11 Sentencing Differentials in Prosecutors' Decisions

	Nonzero Statutory Minimum			Statutory Minimum		
	Unconditional (1)	Conditional (2)	Conditional (3)	Unconditional (4)	Conditional (5)	Conditional (6)
Black	.233***	.168***	.051***	28.966***	21.621***	7.806***
	(.016)	(.014)	(.006)	(1.944)	(1.712)	(.892)
Hispanic	.054	.126***	.056***	4.297	13.879***	7.368***
	(.036)	(.022)	(.009)	(3.915)	(2.457)	(1.017)
Sentencing outcome for Whites		.222			22.1	
Offender, legal, and district controls	No	Yes	Yes	No	Yes	Yes
Offense type codes	No	No	Arrest	No	No	Arrest
p-value: Black = Hispanic	.000	.023	.508	.000	.000	.696
Adjusted R^2	.040	.147	.495	.038	.136	.365
Observations	130,216	130,216	68,216	130,216	130,216	68,216

Note.—Ordinary least squares regression estimates are shown in all columns. Standard errors clustered by district are reported in parentheses. The pre-9/11 sample of 130,895 federal cases is used (those that come up for sentencing from October 1, 1998, to September 9, 2001). The dependent variable in cols. 1–3 is a dummy for whether the initial charge filed by prosecutors has an associated mandatory minimum sentence length. The dependent variable in cols. 4–6 is the mandatory minimum sentence length (including zeros for those without a minimum). In cols. 1 and 4 we condition only on defendant group (White, Black, Hispanic). In cols. 2, 3, 5, and 6 the following additional controls are included: fiscal year dummies, offender characteristics (dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this nonmissing-age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a nonmissing-dependents dummy), legal controls (a dummy whether information on the defense counsel is missing and a nonmissing dummy interacted with the type of defense counsel [privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements]), and federal district dummies. In cols. 3 and 6 we additionally control for the primary offense type as measured at the arrest stage. The p-value in each column is for the null that the coefficients on the Black and Hispanic dummy are equal against a two-sided alternative.

*** Significant at 1%.

The results are shown in table 7: (i) Hispanic defendants initially charged before 9/11 are 7.4 percentage points more likely to receive an initial offense that carries a statutory minimum, corresponding to a 22% increase over the pre-9/11 period (an impact statistically different from that for Blacks; $p = .032$); (ii) their statutory minimum sentence is 10.7 months longer; and (iii) there is no evidence that 9/11 impacts prosecutors' initial offense charges filed against Black defendants along either margin ($\hat{\phi}_B = 0$ in cols. 1 and 2). The magnitude of these responses to 9/11 correspond to (i) 60% of the pre-9/11 Hispanic-White gap in the likelihood of an initial offense charge with a mandatory minimum and (ii) 77% of the pre-9/11 Hispanic-White gap in the statutory minimum sentence length. Indeed, these impacts of 9/11 leave the overall post-9/11 Hispanic-White differential on each margin to be at least as large as the Black-White differential.

This pattern of results closely mirrors those found earlier for judges: they are consistent with the structure of social preferences across out-groups for prosecutors being such that there is contagious animosity from Muslims to Hispanics, while their social preferences are independent between Muslims and Blacks.

In the appendix, we consider two further dimensions of prosecutor behavior: (i) granting of substantial assistance departures (which can occur at the plea stage of the timeline) and (ii) drafting of the presentence report (which occurs between trial and sentencing). For the first dimension, we find no differential impacts on the likelihood that prosecutors grant substantial assistance departures; this helps rule out that the increase in statutory minimum sentence lengths driven by initial offense charges is later undone through defendant cooperation in plea bargains. For the second dimension, for both out-groups we see no change in the minimum sentence in the guideline cell defendants are placed in. Hence prosecutor–legal counsel interactions at the presentence report stage between trial and sentencing are not a major source of differential treatment of defendants by out-group after 9/11. This suggests that increased Hispanic-White sentencing gaps after 9/11 are not due to diminished effort on the part of legal counsel of Hispanic defendants.

VI. Interpretation

We have documented an impact of 9/11 on outcomes for a major (non-Muslim) minority group in the high-stakes and professional environment of the federal CJS. One interpretation is that the changes in behavior of in-group judges and prosecutors are driven by their social preference structures over out-groups. In particular, their behavior can be rationalized by them having contagious animosity from Muslims to Hispanics, while social preferences are independent between Muslims and Blacks. We now probe the data further using two very different approaches to rule out alternative interpretations of $\hat{\phi}_e$.

Table 7
Prosecutors' Initial Charges around 9/11

	Nonzero Statutory Minimum (1)	Statutory Minimum Length (2)
Initial charges post-9/11 × Hispanic	.074*	10.7*
	(.043)	(5.53)
Initial charges post-9/11 × Black	−.010	.684
	(.047)	(7.74)
Initial charges post-9/11	−.033	−5.96
	(.035)	(4.07)
Offender, legal, and district controls	Yes	Yes
Offense type codes	No	No
Guideline cell dummies	No	No
p-value: post-9/11 × Black = post-9/11 × Hispanic	.032	.160
Adjusted R^2	.170	.147
Observations	3,600	3,600

NOTE.—Ordinary least squares regression estimates are shown in all columns. Standard errors clustered by district are reported in parentheses. The sample of federal cases used is as follows: (i) for those with initial charges after 9/11, defendants in (out of) custody committed their last offense between 14 (21) days before 9/11 and the day before 9/11; (ii) for those with initial charges before 9/11, defendants in (out of) custody committed their last offense between 42 (63) days before 9/11 and 38 (42) days before 9/11. The dependent variable in col. 1 is a dummy for whether the defendant receives an initial charge with a nonzero statutory minimum sentence. The dependent variable in col. 2 is the length of statutory minimum sentence. In all columns the following controls are included: offender characteristics (dummies for the highest education level, marital status, a dummy for whether age is missing, age and age squared interacted with this nonmissing-age dummy, a dummy for whether the number of dependents is missing, and the number of dependents interacted with a nonmissing-dependents dummy), legal controls (a dummy whether information on the defense counsel is missing and a nonmissing dummy interacted with the type of defense counsel [privately retained, court appointed, federal public defender, self-represented, rights waived, other arrangements], and federal district dummies. The p-value in each column is for the null that the coefficients on the post-9/11 × Black and post-9/11 × Hispanic dummy interactions are equal against a two-sided alternative.
* Significant at 10%.

A. Decomposition Analysis

We first present a decomposition of sentencing differentials to understand whether they are being driven by changes in observables or sentencing penalties for those observables. We focus on defendants who come up for judicial sentencing just around 9/11, among whom we have documented that Hispanics are significantly less likely to be downward departed (table 3). We use the Juhn, Murphy, and Pierce (1993) decomposition. This is implemented by first considering the following sentencing equation for White defendant i sentenced in period T: $s_{iT}^{\mathrm{W}} = X'_{iT}\beta_T^{\mathrm{W}} + u_T^{\mathrm{W}}\theta_{iT} = X'_{iT}\beta_T^{\mathrm{W}} + \varepsilon_{iT}^{\mathrm{W}}$, where β_T^{W} are sentence penalties for Whites and $\varepsilon_{iT}^{\mathrm{W}}$ is a residual for White defendant i in period T. The explicit assumption is that the residuals and covariates are independent (Fortin, Lemieux, and Firpo 2011). The Hispanic-White sentencing differential in period T is then $\Delta s_T = s_T^{\mathrm{H}} - s_T^{\mathrm{W}} = \Delta X_T\beta_T^{\mathrm{W}} + \Delta\varepsilon_T$. Given our difference-in-differences research design, we take a second difference over pre- to post-9/11 time periods ($T = 0$ to $T = 1$):[18]

$$\Delta s_1 - \Delta s_0 = (\Delta X_1 - \Delta X_0)\beta_0^W + \Delta X_1\left(\beta_1^W - \beta_0^W\right) + (\Delta\varepsilon_1 - \Delta\varepsilon_0)\,. \quad (3)$$

The unconditional difference-in-differences in the likelihood of downward departure to be explained is $\Delta s_1 - \Delta s_0 = -.041$. The $(\Delta X_1 - \Delta X_0)\beta_0^W$ component, or X effect, measures the contribution to the difference-in-differences in sentencing gaps of observables. The $\Delta X_1(\beta_1^W - \beta_0^W)$ component, or β effect, measures changes in sentencing penalties before and after 9/11 for observables.[19]

Figure 6 shows the X and β effects for specific covariates, where the Y-axis shows the implied sentencing differential that can be attributed to each X and β effect. As expected, this shows that each X effect, on quantities, is small. This is because of our research design, and this result is essentially analogous to what was shown in table 2—that defendant observables are balanced before and after 9/11 by group. A more interesting pattern of changing penalties across covariates emerges, with the penalties on some covariates rising and others falling. Because of the alternating signs of the effects, only 7% of the unconditional difference-in-differences is overall attributable to observables either through the X effects or the β effects.

For example, penalties related to education, being married, and having children all rise, suggesting that after 9/11 Hispanics would have been more likely to be downward departed than Whites. On covariates related to offense types, we note that the X and β effects never explain more than 17% of the observed sentencing gap between Hispanics and Whites, while differences in defense counsel types do not explain more than 9% of the overall gap.

Taken together, these findings help rule out explanations for the results based on the harshness with which certain offense types are dealt with after 9/11 and on offender characteristics, including those that might perhaps closely predict recidivism, such as the guideline cell they are assigned to, as well as explanations related to effort or allocation of legal counsel to defendants after 9/11. All of this suggests that explanations for why Hispanic-White sentencing differentials worsen after 9/11 based on statistical

[18] While it is well understood that such decompositions do not represent formal tests for statistical discrimination (Charles and Guryan 2011), in our setting the usual concerns related to decomposition analysis for studying discrimination are partly ameliorated because (i) the difference-in-differences setup provides common support in the cross section of covariates across groups and (ii) the inclusion of guideline cell dummies allows us to capture many case-specific factors driving outcomes.

[19] To check the validity of basing the JMP decomposition on a linear probability model, we have also conducted cross-sectional decompositions in the pre- and post-9/11 periods separately, using a Blinder-Oaxaca decomposition and the Fairlie (2005) extension of such decompositions to nonlinear models. Constructing the implied difference-in-difference decomposition from either approach generates very similar conclusions as the JMP decomposition based on the linear probability model.

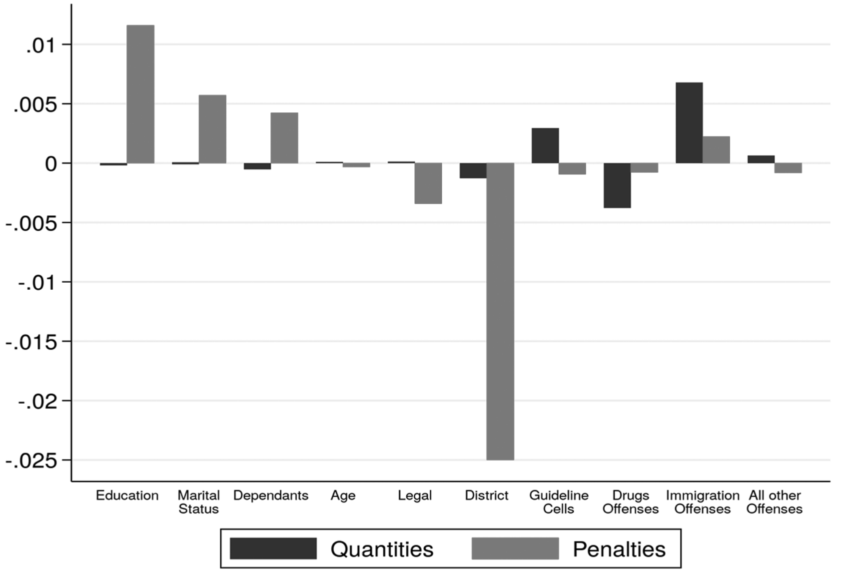

FIG. 6.—Decomposition of Hispanic-White differentials in downward departures. The graph shows key results from a Juhn-Murphy-Pierce (1993) decomposition, using a nonparametric procedure. This decomposes the unconditional difference-in-differences for downward departures between Hispanics and Whites, based on federal criminal cases in the natural experiment sample. Hence, the decomposition is based on 29,352 cases for Hispanic or White defendants that come up for sentencing in a 6-month window on either side of September 11, 2001. For those defendants sentenced after September 11, 2001, the last offense was committed before September 11, 2001, and if sentenced before September 11, 2001, the last offense was committed at least 180 days before September 11, 2001. The controls in this decomposition are offender characteristics, defense counsel type, primary offense type dummies and a dummy for multiple offenses, guideline cell dummies, and federal district dummies. For the Juhn-Murphy-Pierce decomposition, Whites are chosen as the reference group. A color version of this figure is available online.

discrimination alone are not easily reconcilable with the evidence. This is also fits with evidence that recidivism rates did not change across groups before and after 9/11 (BJS 2014, 2018).[20]

[20] BJS (2014) reports recidivism rates by race for two cohorts of defendants: those released in 1994 and those released in 2004. This suggests that (i) 3-year recidivism rates of all groups have risen over time and (ii) there has been no great differential increases across groups over time in recidivism rates. BJS (2018) reports recidivism rates by race over a 9-year follow-up period for defendants released in 2005: this shows that Hispanics have higher 1-year recidivism rates than Whites. However, 9 years after release recidivism rates are found to be almost equal between Whites and Hispanics, but they are higher for Black defendants. In sum, this evidence does not strongly suggest that after 9/11 recidivism rates among Hispanics rose more than those for other groups.

However, the one covariate that can potentially explain the observed sentencing gap is the federal district of the case: the X effect is again small and corresponds only to 3% of the unconditional difference-in-differences, but the β effect can explain 60% of the gap ($-.025$ of the actual gap, $-.041$). We therefore next examine one important source of spatial variation that might be being reflected in increasing penalties in the decomposition: judge characteristics.

B. Judge Characteristics

In federal court data, judge identifiers are typically unavailable (or only a subset of cases can be linked) because these cases are considered more serious and often of national importance.[21] To make progress on how judge characteristics correlate to the change in sentencing differentials, we coded the characteristics of federal judges by district, sourced from the *Biographical Directory of Federal Judges*. This details the race/ethnicity, gender, and seniority of judges in 90 districts and whether they were appointed under a Democrat or Republican president. As described in the appendix, we use this to construct judge characteristics by district ($\mathbf{J}_d$).

Similar to Charles and Guryan (2011), we proceed in two steps. First, we estimate equation (1) allowing for a full set of interactions between each federal district d and ($Hispanic_e \times Post_t$) to estimate the coefficient of interest: $\phi_{\mathrm{H},d}$. We do so for the likelihood of a downward departure. Figure A4 shows the spatial pattern of changes in sentencing differentials, plotting $\hat{\phi}_{\mathrm{H},d}$ for each district. Second, we regress $\hat{\phi}_{\mathrm{H},d}$ against $\mathbf{J}_d$ and other district characteristics, where observations are weighted by the share of defendants in district d in the working sample that are Hispanic. Observations are weighted because the underlying regression from which each $\hat{\phi}_{\mathrm{H},d}$ is estimated is based on individual observations, and these vary by district. Robust standard errors are reported.

The weighted mean share of Hispanic (Black) judges in a district is 14% (7%). We note that 16 of 90 districts (18%) have at least one Hispanic judge, the weighted mean share of Hispanic judges is 13.4%, the median share is 16%, and the share conditional on there being at least one Hispanic judge is 19%. Hispanic judges are more likely to be in districts with more Hispanic defendants: the correlation between the share of Hispanic judges and Hispanic defendants in districts is 0.78 (when districts are weighted by the share of Hispanic defendants). Seventeen percent of judges are women, 28% are of senior status, and 48% are appointed by Democrat presidents. As there are only on average 7.5 judges per district, small changes in the composition of judges can significantly alter a defendant's probability to be sentenced by a minority judge.

Table 8 shows the second-stage results. In column 1 we control only for judge race/ethnicity. We find that in districts where there are a higher proportion of Hispanic judges, the change in the Hispanic-White sentencing

[21] An important relevant exception is Yang (2015), who links individual judge data to federal cases to examine how racial sentencing differentials are impacted once sentencing guidelines were struck down in *United States v. Booker* in 2005.

Table 8
Judge Characteristics (Dependent Variable: Coefficient on Post-9/11 × Hispanic × District Dummy)

	Race/ Ethnicity (1)	Other Judge Characteristics (2)	District Population (3)	Effect Size (4)
District proportion Hispanic judges	.225***	.204**	.554***	.032***
	(.073)	(.101)	(.207)	(.012)
District proportion Black judges	.272	.338	.097	.008
	(.217)	(.222)	(.207)	(.018)
District proportion senior status judges		−.066	.027	.004
		(.076)	(.090)	(.014)
District proportion male judges		−.022	−.143	−.017
		(.095)	(.093)	(.011)
District mean judge age		.006*	.004	.015
		(.003)	(.003)	(.014)
District proportion Democrat president appointees		.180**	.137**	.025**
		(.076)	(.066)	(.012)
District proportion of postperiod window with Bush-appointed US attorney		.026	−.046	−.017
		(.027)	(.033)	(.013)
District proportion Black (2000)			.275**	.032**
			(.127)	(.015)
District proportion Hispanic (2000)			−.337*	−.034*
			(.184)	(.019)
Change in district proportion Black (1990–2000)			−2.59**	−.027**
			(1.06)	(.011)
Change in district proportion Hispanic (1990–2000)			−.100	−.002
			(.519)	(.011)
Mean of dependent variable		−.016		
Adjusted R^2	.105	.172	.287	.287
Observations	88	88	88	88

NOTE.—The results are based on the natural experiment sample (those who come up for sentencing in a 6-month window either side of September 11, 2001, where for those defendants sentenced after September 11, 2001, the last offense was committed before September 11, 2001, and if sentenced before September 11, 2001, the last offense was committed at least 180 days before September 11, 2001). Each observation represents a single federal court district, and observations are weighted by the share of Hispanics in the district in the relevant sample of federal criminal cases (the natural experiment or full sample). Robust standard errors are reported. The dependent variable is the coefficient on post-9/11 × Hispanic × district from a difference-in-difference-in-differences regression for the natural experiment sample period where in this first stage the full set of controls is included and the dependent variable is whether a downward departure is granted. The data for judicial characteristics are sourced from the *Biographical Directory of Federal Judges*. To select the relevant judges to construct characteristics for, we used the data on commission and termination dates for each judge in the database, and we restrict the sample to judges commissioned before the end of the natural experiment sample and those who terminated the bench after the beginning of the sample. Data for district-level population characteristics are from the 1990 and 2000 5% US census county-level data. District proportions were aggregated up from county-level data. In column 4, effect sizes on all covariates are reported.

* Significant at 10%.
** Significant at 5%.
*** Significant at 1%.

differential, $\hat{\phi}_{\mathrm{H},d}$, is significantly smaller. Column 2 shows that this is robust to controlling for the seniority, gender, age, and appointment characteristics of federal district judges as well as the share of the post-9/11 window the district spends under a Bush-appointed US attorney. This suggests that the Hispanic judge effect is not merely capturing them being Democrat appointees, and consistent with the evidence in Schanzenbach (2005) and Harris and Sen (2019), the presence of Democrat-appointed judges has an independent correlation with changes in the Hispanic-White sentencing differential.[22]

Column 3 controls for the population shares of ethnic groups in the district as well the change (1990–2000) in proportions for each group. This increases the coefficient on the district proportion of Hispanic judges from 0.200 to 0.548 (where both are significant at conventional levels), and this partial correlation becomes more precisely estimated. Hence, the district proportion of Hispanic judges does not appear to proxy population characteristics of where the case is heard.

To more easily compare across covariates, column 4 reports effect size estimates of each partial correlation. We see that a 1 standard deviation increase in the proportion of judges in the district of Hispanic origin increases $\hat{\phi}_{\mathrm{H},d}$ by 3.2 percentage points. This effect size is larger than the implied impact on the change in the Hispanic-White sentencing differential of a 1 standard deviation increase in the share of Democrat-appointed judges. The effect size is comparable in absolute magnitude to the average effect across all districts, documented in table 3, that after 9/11, Hispanic defendants are 3.8 percentage points less likely to receive a downward departure.

That judge ethnicity correlates to the change in the Hispanic-White sentencing differential is prima facie evidence against the results being explained by statistical discrimination: if so, then all judges, irrespective of their own characteristics, should use defendant ethnicity as a marker for unobservable traits in determining sentencing outcomes. This is in the spirit of rank order tests used to distinguish statistical discrimination from animus in the literature using data on police arrests or on individual judges (Anwar and Fang 2006; Park 2017).[23] This interpretation is further reinforced by noting that the more experienced judges are uncorrelated with smaller changes in sentencing differentials (measured through either the senior status of judges or their age). This is counter to the Altonji and Pierret (2001) test of statistical discrimination, exploiting the fact that with experience, decision makers learn

[22] Our results are consistent with those of Cohen and Yang (2019), who use individual judge data to show how Republican judges give harsher sentences to Black defendants.

[23] Such hit-rate tests for racial bias in the context of arrest data have been devised to deal with the nonrandom selection of individuals into police stops. In our setting, such concerns over the inframarginality problem of detecting bias are weaker because there is random matching of defendants to judges in the federal CJS.

the true characteristics of agents and become less reliant on proxies such as race/ethnicity.

VII. Conclusions

In-group bias is a central aspect of human behavior, where individuals aid members of a group they socially identify with more than members of other groups they do not identify with as strongly (Tajfel et al. 1971). We extend this notion to contexts in which social preferences are defined over multiple out-groups. We use a quasi-experimental research design around 9/11 to shed new light on the structure of social preferences across out-groups. Our research design allows us to investigate whether increased animosity toward Muslims in the aftermath of 9/11 had spillover effects on Black and Hispanic individuals in the context of the high-stakes and professional environment of the federal CJS.

Our core finding is that as 9/11 sparked a rise in animosity toward Muslims, Hispanic defendants experience worsening sentence and presentence outcomes, in line with judges and prosecutors having social preferences characterized by contagious animosity from Muslims to Hispanics. In contrast, the social preferences of judges and prosecutors are independent between Muslim and Black defendants. We underpin a causal interpretation of these findings by providing evidence in favor of the identifying assumptions underlying our research design, and we narrow down the interpretation of the results by ruling out that they are driven by citizenship or by statistical discrimination against Hispanic defendants. As such, our analysis helps address an appeal made in recent overviews of the economics of discrimination literature on the need to better bridge to the psychology literature on the origins of discriminatory behavior (Charles and Guryan 2011; Bertrand and Duflo 2017). We do so with two important caveats: (i) we have exploited a particularly traumatic event that could have triggered a strong emotional response, even in this high-stakes setting, in line with nascent well-identified causal evidence on emotions driving judicial decisions (Shayo and Zussman 2011; Chen, Moskowitz, and Shue 2016; Philippe and Ouss 2018); and (ii) our research design does not allow us to estimate whether the impacts persist beyond the short-run window of cases in our sample.

Our findings provide among the first field evidence of contagious animosity, that is, that social preferences across out-groups are malleable. This adds to a nascent body of work examining the structure of social preferences, which has so far typically been based on self-reported or observational data collected in postconflict environments (Bauer et al. 2016). An important implication of our findings is that antidiscrimination policies toward one out-group can have externalities on other out-groups. On policy implications, our results suggest that appointing more Hispanic judges to federal district courts or as federal prosecutors might go some way toward reducing Hispanic-White sentencing differentials.

Two directions for future research are clear. First, in keeping with the earlier literature on in-group bias, we do not estimate the extent to which in-group members have heterogeneous preferences toward out-groups, and so it is as if we assume homogeneity of preferences within groups. As judges are randomly assigned, our estimates reflect average sentencing differentials driven by the behavior of judges and prosecutors. This is in contrast to what is observed in labor market studies of discrimination: one of Gary Becker's key insights was that observed racial wage gaps do not reflect average levels of employer discrimination because minority employees can sort toward the least discriminating employer. If there is a sufficiently large share of minority workers relative to nondiscriminating employers, the equilibrium wage gap reflects the tastes of the marginal employer. In our context, the lack of defendant-judge sorting is what leads us to measure average levels of animus.

Yet there is clearly much work to be done to understand within-group heterogeneity and correlates of idiosyncratic variation in social preference structures within groups. A promising avenue in this context is to build on Yang (2015) and link individual judge data to federal cases for our sample period. Utilizing such information would help shed light on individual characteristics correlated with the structure of social preferences and so might have implications for how sentencing disparities could be mitigated through the initial selection or training of federal judges.

In addition, there are many potential out-groups one could consider for which there is a rich set of social preference structures to identify. There is no reason to expect contagious animosity/altruism to characterize all pairs. More broadly, there can be circumstances in which individuals have multiple identities, and there are other circumstances in which individuals can endogenously choose an identity in anticipation of the kinds of interlinked social preference structures we have documented. This opens up a wide array of research questions at the intersection of the formation of social preferences and the economics of identity.

References

Abrams, David, Marianne Bertrand, and Sendhil Mullainathan. 2012. Do judges vary in their treatment of race? *Journal of Legal Studies* 41:347–83.

Alexander, Richard D. 1987. *The biology of moral systems*. New York: Aldine De Gruyter.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113:151–84.

Altonji, Joseph G., and Charles R. Pierret. 2001. Employer learning and statistical discrimination. *Quarterly Journal of Economics* 116:313–50.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson. 2012. The impact of jury race in criminal trials. *Quarterly Journal of Economics* 127:1017–55.
Anwar, Shamena, and Hanming Fang. 2006. An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review* 96:127–51.
Bauer, Michal, Christopher Blattman, Julie Chytilová, Joseph Henrich, Edward Miguel, and Tamar Mitts. 2016. Can war foster cooperation? *Journal of Economic Perspectives* 30:249–74.
Bertrand, Marianne, and Esther Duflo. 2017. Field experiments on discrimination. In *Handbook of field experiments*, ed. Abhijit Banerjee and Esther Duflo. Amsterdam: North-Holland.
BJS (Bureau of Justice Statistics). 2014. Recidivism of prisoners released in 30 states in 2005: Patterns from 2005 to 2010. NCJ 244205.
———. 2018. Update on prisoner recidivism: A 9-year follow-up period 2005–2014. NCJ 250975.
Boyd, Robert, Herbert Gintis, Samuel Bowles, and Peter J. Richerson. 2003. The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the USA* 100:3531–35.
Bushway, Shawn D., and Anne M. Piehl. 2001. Judging judicial discretion: Legal factors and racial discrimination in sentencing. *Law and Society Review* 35:733–64.
Charles, Kerwin K., and Jonathan Guryan. 2011. Studying discrimination: Fundamental challenges and recent progress. *Annual Review of Economics* 3:479–511.
Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue. 2016. Decision making under the gambler's fallacy: Evidence from asylum judges, loan officers, and baseball umpires. *Quarterly Journal of Economics* 131:1181–242.
Cohen, Alma, and Crystal Yang. 2019. Judicial politics and sentencing decisions. *American Economic Journal: Economic Policy* 11:160–91.
CRS (Congressional Research Service). 2013. The federal prison population buildup: Overview, policy changes, issues and options. Report 7-5700. Washington, DC: CRS.
Davis, Darren 2007. Negative liberty: Public opinion and the terrorist attacks on America. New York: Russell Sage Foundation.
Eifert, Ben, Edward Miguel, and Daniel N. Posner. 2010. Political competition and ethnic identification in Africa. *American Journal of Political Science* 54:494–510.
Fairlie, Robert W. 2005. An extension of the Blinder-Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement* 30:305–16.
Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2007. Unconditional quantile regressions. *Econometrica* 77:953–73.
Fischman, Joshua B., and Max M. Schanzenbach. 2012. Racial disparities under the federal sentencing guidelines: The role of judicial discretion

and mandatory minimums. *Journal of Empirical Legal Studies* 9:729–64.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2011. Decomposition methods in economics. In *Handbook of labor economics*, vol. 4A, ed. Orley C. Ashenfelter and David Card. Amsterdam: Elsevier.

Fowler, James H., and Nicholas A. Christakis. 2010. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences of the USA* 107:5334–38.

Glaeser, Edward L., Daniel P. Kessler, and Anne M. Piehl. 2000. What do prosecutors maximize? An analysis of the federalization of drug crimes. *American Law and Economics Review* 2:259–90.

Gould, Eric D., and Esteban F. Klor. 2016. The long-run effect of 9/11: Terrorism, backlash, and the assimilation of Muslim immigrants in the West. *Economic Journal* 126:2064–114.

Harris, Allison P., and Maya Sen. 2019. Bias and judging. *Annual Review of Political Science* 22:241–59.

Hopkins, Daniel J. 2010. Politicized places: Explaining where and when immigrants provoke local opposition. *American Political Science Review* 104:40–60.

Human Rights Watch. 2002. We are not the enemy: Hate crimes against Arabs, Muslims, and those perceived to be Arab or Muslim after September 11. Human Rights Watch 6.

Jeffries, John C., Jr., and John Gleeson. 1995. The federalization of organized crime: Advantages of federal prosecution. *Hastings Journal* 46:1095–134.

Jordan, Jillian J., David G. Rand, Samuel Arbesman, James H. Fowler, and Nicholas A. Christakis. 2013. Contagion of cooperation in static and fluid social networks. *PLoS ONE* 8:e66199.

Juhn, Chinhui, Kevin M. Murphy, and Brooks Pierce. 1993. Wage inequality and the rise in returns to skill. *Journal of Political Economy* 101:410–42.

Klepper, Steven, Daniel Nagin, and Luke-Lon Tierney. 1983. Discrimination in the criminal justice system: A critical appraisal of the literature. In Research on Sentencing: The Search for Reform, vol. 2, ed. A. Blumstein, J. Cohen, S. E. Martin, and M. H. Tonry. Washington, DC: National Academy Press.

Legewie, Joscha. 2013. Terrorist events and attitudes towards immigrants: A natural experiment. *American Journal of Sociology* 118:1195–245.

Liptak, Adam. 2003. For jailed immigrants, a presumption of guilt. *New York Times*, June 3.

Mueller-Smith, Michael. 2016. The criminal and labor market impacts of incarceration. Unpublished manuscript, University of Michigan.

Mustard, David B. 2001. Racial, ethnic and gender disparities in sentencing: Evidence from the US federal courts. *Journal of Law and Economics* 44:285–314.

Orrenius, Pia M., and Madeline Zavodny. 2009. The effects of tougher enforcement on the job prospects of recent Latin American immigrants. *Journal of Policy Analysis and Management* 28:239–57.

Oster, Emily. 2019. Unobservable selection and coefficient stability: Theory and validation. *Journal of Business Economics and Statistics* 37:187–204.

Park, Kyung H. 2017. Do judges have tastes for discrimination? Evidence from criminal courts. *Review of Economics and Statistics* 99:810–23.

Philippe, Arnuad, and Aurélie Ouss. 2018. "No hatred or malice, fear or affection": Media and sentencing. *Journal of Political Economy* 126:2134–78.

Rehavi, Marit M., and Sonja B. Starr. 2014. Racial disparity in federal criminal sentences. *Journal of Political Economy* 122:1320–54.

Romero, Luis A., and Armino Zarrugh. 2018. Islamophobia and the making of Latinos as terrorist threats. *Ethnic and Racial Studies* 41:2235–54.

Schanzenbach, Max M. 2005. Racial and sex disparities in prison sentences: The effect of district-level judicial demographics. *Journal of Legal Studies* 34:57–92.

Schanzenbach, Max M., and Emerson H. Tiller. 2007. Strategic judging under the US sentencing guidelines: Positive political theory and evidence. *Journal of Law, Economics, and Organization* 23:24–56.

Shayo, Moses. 2009. A model of social identity with an application to political economy: Nation, class and redistribution. *American Political Science Review* 103:147–74.

Shayo, Moses, and Asaf Zussman. 2011. Judicial ingroup bias in the shadow of terrorism. *Quarterly Journal of Economics* 126:1447–84.

Starr, Sonja B., and Marit M. Rehavi. 2013. Mandatory sentencing and racial disparity: Assessing the role of prosecutors and the effects of *Booker*. *Yale Law Journal* 123:2–80.

Suri, Siddharth, and Duncan J. Watts. 2011. Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE* 6:e16836.

Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament. 1971. Social categorization and intergroup behavior. *European Journal of Social Psychology* 1:149–78.

USSC (US Sentencing Commission). 1999–2003. Monitoring of Federal Criminal Sentences, 1999–2003 [computer file], ICPSR version. Washington, DC: USSC (producer), 1999–2006; Ann Arbor, MI: ICPSR (distributor).

Woods, Joshua. 2011. The 9/11 effect: Toward a social science of the terrorist threat. *Social Science Journal* 48:213–33.

Yang, Crystal S. 2015. Free at last? Judicial discretion and racial disparities in federal sentencing. *Journal of Legal Studies* 44:75–111.

———. 2016. Resource Constraints and the criminal justice system: Evidence from judicial vacancies. *American Economic Journal: Economic Policy* 8:289–332.