A Multidimensional Examination of Jury Composition, Trial Outcomes, and Attorney Preferences*

Jee-Yeon K. Lehmann† University of Houston

Jeremy Blair Smith‡ Boston University

June 27, 2013

Abstract

We assess the degree to which seated juries in U.S. criminal trials might fall short of the constitutional ideal of impartiality. We first ask if certain demographic and socioeconomic characteristics are related to pre-deliberation biases that individual jurors hold or to the verdicts at which juries arrive collectively. We do not focus solely on race, but also jointly consider other characteristics – sex, age, religiousness, education, and income – that existing literature has largely neglected. A uniquely rich dataset on non-capital felony jury trials held in four major state trial courts allows us to identify within-case effects and to control for typically unobservable aspects of the trial and its participants. We find that jurors with higher income and religiousness hold more favorable sentiments for the prosecution, while blacks hold more favorable sentiments for the defense. These pre-deliberation biases are reflected in trial outcomes, with juries with a higher average income and a greater proportion of religious jurors acquitting on fewer counts, and juries with a greater proportion of blacks convicting on fewer counts. However, these jury composition effects are smaller and account for less of the explained variation in verdicts than the effect of evidentiary strength. Moreover, while we find that prosecuting/defense attorneys prefer juries with higher/lower average income, indicating that attorneys are aware of the effect of income on predispositions and verdicts and may therefore attempt to leverage this knowledge to manipulate trial outcomes in their favor, we also find evidence that they are mistaken about the effects of other characteristics. Our results thus raise some concerns regarding the trustworthiness of U.S. criminal trials, but also provide important context for such concerns, especially by illustrating that the sources of jury bias may be more nuanced and multidimensional than an analysis based on race alone would imply.

JEL Codes: K4, J7

*We are especially grateful to Kevin Lang, Claudia Olivetti, and Michael Manove for their guidance and encouragement on this project. Special thanks to Johannes Schmieder, Marc Rysman, Daniele Paserman, Aimee Chin, Steven Craig, Scott Imberman, Chinhui Juhn, Jungmo Yoon, Andy Zuppann, Michael Luca, David Rossman, and Hsueh-Ling Huynh for helpful comments. Thanks as well to the many other colleagues and seminar participants who have offered feedback along the way. Any errors are our own.

†E-mail: jlehmann@uh.edu. This material is partly based upon graduate work supported under the Department of Education Jacob K. Javits Fellowship, the Mustard Seed Foundation Harvey Fellowship, and the Boston University Tuition Fellowship.

‡E-mail: jersmith@bu.edu. I would like to acknowledge financial support from the Social Sciences and Humanities Research Council of Canada and the Boston University Department of Economics.

1 Introduction

Legal scholars and the public have frequently voiced concerns that the American legal system fails to provide defendants with their constitutional right to an impartial jury in criminal prosecutions.[1] The focal point of such concerns is typically the jury selection process, which is regarded as giving attorneys a means to stack juries with biased jurors.[2] But underlying this view is a supposition that attorneys can identify such biased jurors in the first place. This identification is straightforward if juror biases are correlated with observable juror characteristics such as race, which they are commonly assumed to be. Thus, the trial of the police officers charged with the beating of Rodney King – in which an entirely non-black jury failed to convict four white men of excessive force against a black victim despite visceral video evidence – is seen as epitomizing a modern manifestation of the old maxim attributed to Clarence Darrow that a case is won or lost by the time the jury is sworn in.[3]

In this paper, we assess the degree to which such concerns are justified along several dimensions. First, we ask if certain demographic and socioeconomic characteristics are indeed related to the sentiments and biases that individual jurors take with them into the deliberation room or to the verdicts at which juries collectively arrive. In examining this question, we do not focus solely on race, but also jointly consider other juror characteristics – such as sex, age, religiousness, education, and income – that existing literature has largely neglected. Second, we place the size of the jury composition effects on verdicts in the context of other important aspects of the trial, especially the strength of the evidence presented in the courtroom. Finally, we assess whether attorneys are aware of the effects of juror characteristics on pre-deliberation biases and trial outcomes. In addressing these questions, we take advantage of a uniquely rich dataset on non-capital felony jury trials held in four major U.S. state trial courts[4] during 2000 to 2001 that includes detailed survey information on various aspects of the case from all trial participants, including the jurors, the judge, and the attorneys.

We find strong evidence of significant relations between individual jurors' demographic and socioeconomic characteristics and their pre-deliberation leanings. In particular, we find

[1] The Sixth Amendment of the U.S. Constitution states that "The accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed."

[2] For example, Rothwax (1996) suggests that attorneys "can mold a jury in the hope that it will be swayed by emotion and innuendo, not fact."

[3] In the initial state trial of the officers, the jury was comprised of ten Caucasians, one white Hispanic, and one Asian. Three of the officers were acquitted of all charges, while the jury was hung on one charge and acquitted on others for the fourth officer. The infamous videotape of the beating was, of course, not the only evidence presented at trial, and it remains difficult to predict if the trial outcome would have been different if a jury of a different racial composition had heard all of the same evidence, testimony, and judge instructions. See Linder (2007) for further details and discussion.

[4] Los Angeles (CA), Maricopa (AZ), Bronx (NY), and Washington D.C.

that jurors with higher levels of education and religiousness and, most consistently, those with higher income are significantly more likely to hold biases and sentiments favoring the prosecution compared to other jurors who heard the same case.[5] Female jurors hold some pre-deliberation opinions weakly favoring the prosecution that depend to some degree on the gender of the victim. Black jurors show greater favoritism for the defense in general, and especially tend to express pro-defense sentiments when the defendant is also black.

We then examine whether the juror characteristics that are associated with significant individual pre-deliberation biases also have strong effects on the verdicts arrived at post deliberation. Using a rich set of case-level controls, including attorney demographics and experience measures as well as the judge's assessment of attorney skills and trial evidence, we find that juries with higher average income and religiousness hand out significantly fewer acquittals, closely mirroring our findings regarding jurors' pre-deliberation biases. We also find that juries with a greater proportion of blacks convict on fewer counts, predominantly when both the defendant and victim are black. Finally, we find that a greater proportion of females and greater average age and years of education amongst seated jurors are weakly associated with trial outcomes favoring the defendant. These jury composition effects are small compared to the effect of evidentiary strength, and account for less of the explained variation in verdicts than do the factual and legal aspects of the trial. Specifically, reducing acquittals by the same amount as would be induced by increasing the strength of the evidence by one standard deviation would require, other things equal, replacing four black jurors with non-black jurors on a twelve-person jury or increasing the income of each juror by about 40%.

Having established the significance of jury characteristics in shaping juror predispositions and trial outcomes, we examine whether attorneys correctly anticipate these associations by asking if they hold preferences for juror characteristics that tend to favor their side. Capitalizing on the availability of a self-reported measure of attorney satisfaction regarding the seated jury, we ask whether the prosecutor's and the defense attorney's levels of satisfaction can be explained by the characteristics of the seated jury, once again conditioning our analysis on a rich set of controls.[6] Robust to all specifications, we find that seated juries with higher average income are associated with significantly higher satisfaction levels for the prosecution and lower satisfaction levels for the defense. This is fully consistent with our findings that higher income is strongly related to pro-prosecution juror sentiments and

[5]The particular dimensions along which we examine juror bias include sympathy for the victim, trust in the police, subjective evaluation of each side's overall case, and interpretation of the trial evidence.

[6]We also limit our analysis to cases in which attorneys reported their satisfaction level with the seated jury *before* learning the verdict, so that their evaluations should not be influenced by the eventual trial outcome.

fewer acquittals. On the other hand, in contrast to the weak association between gender and verdicts and to the tendency for female jurors to hold some biases favoring the prosecution, we find strong evidence that defense attorneys prefer having more women on the jury.

Our results regarding prosecuting and defense attorney preferences for blacks on the seated jury are inconclusive. While the effects of race on attorney satisfaction are sometimes large, they are often very imprecisely estimated. Further, they suggest that prosecuting attorneys may sometimes prefer juries with a greater proportion of blacks, in contrast to our findings that black jurors tend to hold strong predispositions in favor of the defense in general. We speculate on some factors that may be confounding the estimation and interpretation of these race effects on attorney satisfaction, including the fact that it is ostensibly illegal to remove potential jurors on the sole basis of race during jury selection.[7]

Our analysis improves upon existing studies in several ways. First, while past studies of the impacts of jury composition on trial outcomes have focused predominantly on race (e.g. Anwar et al., 2012b; Lee, 2010; Bowers et al., 2001), our results demonstrate that other demographic and socioeconomic characteristics also play an important role in shaping the biases of juries. Furthermore, accounting jointly for race and other characteristics correlated with race allows us to disentangle effects that may have been misattributed solely to race in past studies that omit these other characteristics.[8] Second, our dataset contains trials from multiple courtrooms across several jurisdictions with varying racial diversities and histories of racism. As such, results from this paper should be helpful in assessing the nature and the extent of jury composition effects on trial outcomes, as well as the degree to which attorneys have preferences over certain juror characteristics, outside the deep southern states and across counties with varying racial mixes.[9] Third, we take a stated preference approach to assessing

[7]The Supreme Court ruling in *Batson v. Kentucky* (1986) and other subsequent rulings forbid attorneys to remove jurors solely on the basis of race or gender. However, the effectiveness of these rulings has been widely questioned. This will be discussed in more detail in the following section.

[8]Sommers and Ellsworth (2003) review evidence from the legal literature on the effects of race in criminal jury trials. Also see Pfeifer (1990) for a critical appraisal of some of the earliest of this evidence. Lieberman and Sales (2007) and Baldus et al. (2001) also review evidence on the effects of gender and other socio-demographic characteristics. The studies reviewed by these authors are often based on small samples constructed from archival material, on mock trials, or on public opinion polls, and consider the selected juror characteristics in isolation rather than jointly. In contrast, Anwar et al. (2012b) account for age and gender jointly with race, while Anwar et al. (2012a) undertake a similar joint analysis focusing on age, both employing a large dataset on actual trials. Shayo and Zussman (2011), Alesina and La Ferrara (2011), Abrams et al. (2011), and Iyengar (2011) provide evidence that the race of judges, victims, and defendants can affect trial outcomes in other settings.

[9]Anwar et al. (2012a,b) use data from Lake County and Sarasota County in Florida, which are predominantly white. Rose (1999) uses data from a county in North Carolina described as "largely biracial". Bowers et al. (2001) use data from the Capital Jury Project, which encompasses trials from 14 states, though 11 of these are in the south. Diamond et al. (2009) use data from just a single courtroom, though in the more racially diverse Cook County, Illinois. Baldus et al. (2001) use data from Philadelphia.

which juror characteristics are preferred by attorneys on both sides. Past studies have instead taken a revealed preference approach by exploring the types of potential jurors removed by each attorney, which may be more susceptible to confounding effects of constraints faced by attorneys during jury selection.[10] Finally, due to the richness of our dataset, we are able to incorporate into our analysis many aspects of trials and participants that are unobserved in most previous studies, including, most notably, information on the evidence presented at trial and jurors' interpretation of that evidence.

Overall, our results give some justification for concerns about a lack of impartiality and trustworthiness in U.S. criminal jury trials. One implication is that defendants face the prospect of an unfair trial simply by virtue of being unlucky in the characteristics of the jurors drawn from the jury pool on the day of the trial. But an even more disconcerting implication is that attorneys, with at least some apparent ability to correctly anticipate relations between juror characteristics and biases, might attempt to use this knowledge to manipulate trial outcomes. However, our results also provide some important context for such concerns. First, our multidimensional approach reveals that the sources of jury bias are more nuanced than an analysis based on race alone would suggest. Second, our results on attorney preferences indicate that attorneys have only a partial understanding of these sources of jury bias, which may limit their opportunities to attempt to leverage them to their advantage. Finally, our comparison of the effects of jury composition with that of evidentiary strength suggest that the fundamental issues of fact and law remain the primary determinants of the outcome of any given trial. Investigating these issues further and exploring how our results generalize to regions and types of cases not covered by our dataset are crucial next steps for evaluating the fairness of jury trials and the validity of the verdicts delivered by the U.S. justice system.

The rest of the paper is organized as follows. Section 2 provides a description of the jury selection process and other relevant institutional details. Section 3 describes the dataset and the restricted sample of cases used in our empirical analysis. Section 4 examines the link between juror characteristics and both individual biases and trial verdicts. In Section 5, we present our investigation of attorney preferences over juror characteristics. Finally, we conclude and discuss some additional policy and research questions that our findings raise.

[10]This method has been implemented by Anwar et al. (2012a) to identify attorney preferences over juror age; by Baldus et al. (2001) to identify attorney preferences over juror race, gender, and age; by Grosso and O'Brien (2012) and Rose (1999) to identify attorney preferences over juror race; and, for civil trials, by Diamond et al. (2009) to identify attorney preferences over juror race, gender, age, and income. The jury selection process and the various restrictions that can be imposed upon attorneys during the process will be discussed in the following section.

2 Institutional Background

Throughout our analysis, we will be using data on the characteristics of seated jurors. It is thus important to understand the many forces that can influence which jurors come to be seated on the jury for a given criminal trial.

The broad goal of the overall jury selection process is to randomly draw a panel – also called a venire – of potential jurors from the population of the county in which the alleged crime was committed; then, through a procedure known as voir dire, to remove any potential jurors with an inability to be impartial from the possibility of serving on the seated jury. Of interest here are the details of how the process is actually implemented, and how this broad goal can fail to be met.[11]

There are a number of reasons why the initial panel may not be a random draw from the county population. Juror rolls, from which the names of those to be summonsed on a particular day are randomly drawn, are usually constructed from drivers' license, tax, voter registration, or other administrative databases, and these can systematically under-represent certain groups. Summonses that are undeliverable, never responded to, or received by those with automatic exemptions can likewise over-represent certain groups.[12]

There is also the question of which cases actually make it to a jury trial in a certain county. Defense attorneys are occasionally successful at requesting a change of venue into or out of a given county on the grounds that it would be too difficult to form an impartial jury in the original trial location. Defendants sometimes request a bench (judge-only) rather than a jury trial. And the set of cases that reach a conclusion prior to a jury rendering a verdict or are never prosecuted is determined by a series of decisions and negotiations made by the court and attorneys on both sides. All of these issues raise the possibility that the types of cases heard by a jury in a given courtroom can depend to some extent on attorney expectations of the types of jurors who will eventually comprise the jury.

We will not attempt to diagnose or account for potential non-randomness in the venire or in the trials represented in our dataset. As with past studies, we have no satisfactory means to accomplish this. At the same time, we note that the states covered by our dataset have been above-average or amongst the leaders in jury reform efforts related to jury pool diversity (Mize et al., 2007), and that defense motions for changes of venue are rarely successful outside

[11] Gobert et al. (2009) and Starr and McCormick (2001) each provide a very comprehensive description of jury selection, from both a theoretical and practical perspective. This section draws primarily from these two sources and from a number of interviews we have conducted with legal professionals who have experience with the process.

[12] See Mize et al. (2007), Sommers (2008), and American Civil Liberties Union of Northern California (2010) for further details.

of very high-profile cases.[13] We, therefore, do not expect these issues to substantially affect our results. On the other hand, voir dire presents a serious potential source of endogeneity, which we make a concerted effort to redress in our analysis. Voir dire can be leveraged by attorneys for far more than simply identifying potential jurors with an inability to be impartial. In fact, lawyers are taught to identify potential jurors who are likely to hold unfavorable predispositions to their side and to attempt to remove them. This often takes the form of targeting specific socio-demographic groups that community surveys or mock trials have suggested are more likely than others to hold certain opinions on various aspects of the case.[14] However, attorneys face a number of institutional constraints in this endeavor and must also consider the strategies of their opposing counsel.

During voir dire, potential jurors are questioned, and on the basis of their responses, can be *deselected from* (rather than selected to) the seated jury. The sole legal rationale for voir dire is to ensure that the seated jury is impartial. Individual judges have wide latitude in deciding how this is to be accomplished: there are no constitutionally-mandated procedures, and while most states and some local courts have guidelines or a set of common practices in place, these are rarely binding for any given trial.

The examination of potential jurors can be carried out primarily by the judge, by the judge with suggestions or a greater degree of participation from the attorneys, or primarily by the attorneys. It can also be carried out in the open courtroom or, less often, privately in the judge's chambers or at the sidebar. The primary concerns that are examined in any voir dire are the existence of personal relationships between panelists and any other parties involved in the trial, and the capacity to understand and contemplate the relevant legal issues and evidence in a dispassionate and impartial manner. But many other subjects are often explored as well, especially if attorneys have a high degree of participation. Attorneys may also have access to the results of questionnaires filled in by potential jurors: in almost all cases, they observe the information on the potential jurors that the court collects, which includes basic demographic and occupational data at the least; and they are sometimes also permitted to design and administer "Supplementary Juror Questionnaires" (SJQs) to elicit more detailed information. The judge has discretion over each of these facets of the examination phase; individual judges tend to have strong views on how these decisions should

[13]For example, litigation consultant Gary Moran has stated that such requests are often motivated primarily by delay and diversion. He concludes: "The judge is going to sit there and sit there and after four hours ... he will finally utter the magic word – 'Denied.' And then we all move along" (quoted in Kressel and Kressel, 2002, 59).

[14]In addition to Gobert et al. (2009) and Starr and McCormick (2001), which include much practical instruction for attorneys, see also Hoffman (2006) and Kressel and Kressel (2002) on attorney strategies, and Lieberman and Sales (2007) and Posey and Wrightsman (2005) on the professional practice of what has come to be known as scientific jury selection.

be made, and to hold adamantly to them.

Attorneys can remove panelists via two instruments: strikes for cause and peremptory strikes. An attorney on a given side can challenge a potential juror for cause by arguing that an inability to be impartial has been demonstrated, and the challenge can then be debated amongst the attorneys and the judge. If the judge rules in favor of the strike, the potential juror is dismissed. Attorneys can strike an unlimited number of potential jurors for cause, as long as the judge can be sufficiently convinced that a basis for disqualification has been demonstrated. Each attorney can also remove a limited number of potential jurors without stating any reason by exercising peremptory strikes. The total number of peremptory strikes available is generally small relative to the size of the panel and is usually divided equally between the defense and prosecution, though again, these matters are ultimately at the discretion of the judge.[15]

Although attorneys ostensibly have full discretion over how to exercise their peremptory strikes, they are in fact legally obliged to do so in a non-discriminatory manner. If an attorney suspects that the opposing counsel has exercised a peremptory strike solely on the basis of a potential juror's race or gender, a "Batson objection" can be raised.[16] When a Batson objection is raised and sustained by the judge, the opposing counsel must provide a race- or gender-neutral reason for the strike in order for it to be allowed. If such a reason is not provided, the strike will be disallowed, and other remedies can occasionally be imposed at the judge's discretion as well, even leading to a new voir dire or a mistrial in extreme cases. Many legal commentators regard Batson as only a minor impediment to how attorneys exercise peremptory strikes, since attorneys can and often do comply with its requirements by stating neutral but extremely arbitrary and improbable reasons for the strikes.[17] Less direct ways in which Batson might influence attorney behavior, such as heightening concerns over reputation or the likelihood of verdicts being overturned on appeal, have not been studied.

This description indicates that attorneys have some control over the composition of juries, with the degree of control increasing in the number of peremptory strikes available and

[15]The order in which strikes are made – across attorneys and types of strikes – depends to some extent on whether the struck jury method or some version of the strike-and-replace method is being employed. These are futher details that can vary widely by trial and courtroom, though they are less relevant to the present discussion. See Gobert et al. (2009) and the related discussion in Lehmann and Smith (2012) for more on variations in the sequence in which voir dire can be conducted.

[16]The seminal ruling in *Batson v. Kentucky* (1986) established that the prosecuting attorney must be able to state a race-neutral justification if the defense counsel can make a *prima facie* case that a peremptory strike has been exercised against an African American potential juror solely on the basis of race. Further Supreme Court rulings have extended the applicability of Batson to peremptory strikes by defense as well as prosecuting attorneys, to all races, to discrimination on the basis of gender as well as race, and to civil as well as criminal trials.

[17]See, for example, Equal Justice Initiative (2010) and Norton et al. (2007).

in the freedoms afforded to attorneys in eliciting information about the potential jurors.[18] Attorneys can thus influence trial outcomes through multiple channels: by affecting which jurors from the pool of potential jurors with heterogeneous biases will hear the case; and through the usual processes of argument and presentation of evidence during the trial itself. One would thus not be surprised to find correlations between verdicts and jury composition, since both can be affected by common elements of attorney strategies, which themselves can depend on other aspects of the case. In order to isolate the causal effect of jury composition on trial outcomes, we will thus need to address this endogeneity. We intend to do so by controlling for a rich set of covariates reflecting the freedoms and constraints faced by attorneys in voir dire, the attorneys' skills in designing and implementing effective strategies, and the aspects of each trial that attorneys might condition their strategies on. We will return to this in Section 4, but will first introduce our dataset and the wealth of information it contains.

3 Data Description

3.1 The NCSC/ICPSR Hung Juries Dataset

Our empirical analysis relies on a detailed examination of 351 felony jury trials in four major county courts across the U.S. collected by Hannaford-Agor et al. (2002, 2003) for the National Center for State Courts (NCSC) during 2000 and 2001 and disseminated by the Interuniversity Consortium for Political and Social Research (ICPSR). The four courts – Los Angeles County Superior Court in California, Maricopa County Superior Court in Arizona, Bronx County Supreme Court in New York, and District of Columbia Superior Court in Washington, D.C. – were chosen based on their high volume of felony jury trials and their willingness to cooperate with the data collection process and guidelines. Although the NCSC's main goal in the study was to provide an empirical evaluation of hung juries, the dataset is not limited to trials ending in a hung jury, and indeed includes all non-capital felony trials held in these courts during the specified periods of data collection.[19]

The NCSC study provides a comprehensive look at the trial and all the parties involved: the defendant(s) and the victim(s) (if any), the judge, the attorneys, and, most importantly

[18] We provide some theoretical and empirical support for this intuition in Lehmann and Smith (2012), and also confirm the importance of relative attorney skill. See also Neilson and Winter (2000) and Ford (2010).

[19] Each county had a single period of data collection ranging from 4 to 11 months in length, with data collection ending in Los Angeles in October 2000 before beginning elsewhere, though with the other three periods overlapping in various months in 2001. D.C. contributes the largest number of trials to the dataset, though each court is fairly evenly represented. The death penalty was available in all counties except for D.C. at the time of data collection, though the distribution of charges in the dataset is roughly the same in each county. It should be noted that the confidentiality rules agreed to by the courts precluded any trial participant from accessing any of the survey information at the time of the respective trial or afterwards.

for our analysis, the seated jurors. We do not have information about the initial jury pool from which the seated jury was formed, except for the initial panel size.[20]

The case data in the NCSC study provide researchers with an extensive look at all the charges, the race and sex of the defendant(s) and victim(s), and the voir dire process that led to the seated jury. The NCSC questionnaires also asked the presiding judge of each trial for his/her evaluation of the evidence, case complexities, and attorney skills. The main variables of interest for our empirical analysis come from the attorney and the jury questionnaires. The attorneys in each trial were asked about their satisfaction with the voir dire process, legal experience, and basic demographic information, including sex and age. In addition, the juror questionnaire provides a rich set of demographic information about each seated juror, including age, sex, race, education, income, and religious beliefs. The NCSC data also contain a description of the dynamics of each juror's opinion formation, and of pre- and post-deliberation perceptions and opinions regarding the defendant(s) and victim(s). To our knowledge, the richness of the jury demographic and opinion information available in the NCSC study is unmatched by any other readily available dataset.

Regrettably, although the NCSC dataset contains information on 351 cases and 3,497 jurors in total, not every juror in our data answered every survey question. Since much of our analysis relies on the average characteristics of the seated juries, we restrict our sample of interest to cases in which these average characteristics are well-measured. Specifically, we focus on trials for which six or more jurors responded to questions about education, income, age, race, gender, and religion to reduce any bias that may result in our calculation of average jury characteristics due to systematic non-reporting.[21] Additionally, we further restrict our sample to those cases in which attorneys reported their voir dire satisfaction measures *before* learning the verdict, in order to minimize the possibility that our measure of voir dire satisfaction is confounded by satisfaction with broader aspects of the trial, specifically outcomes.

[20] We can, however, infer the average characteristics of the panel from those of the county as a whole, since, as was discussed previously, a panel is essentially a random sample of the county population. We include county fixed effects in all of our empirical specifications, in part to account for the composition of the panel as much as possible.

[21] Small changes to this cut-off point do not alter out main findings. All counties represented in our dataset were required to have twelve-person juries for felony trials at the time of data collection, except Maricopa, where eight-person juries were permitted for felony trials with charges below a certain level of severity. We do not directly observe which trials in Maricopa actually had eight-person juries, but using a smaller cut-off for all trials that were eligible for eight-person juries does not affect our results.

3.2 Summary Statistics

Table 1 provides some descriptive statistics of our restricted sample. Panel A shows that the majority of cases in our restricted sample fall into the categories of property or drug-related crimes. Consequently, a possible criticism of our dataset is that most of our cases are restricted to non-capital felonies that are not as "serious" as homicides (actual or attempted) or sexual crimes. In turn, attorneys might not pursue the same strategies, and jurors might not act as strongly on their personal prejudices as would be the case with more serious charges. Therefore, the results that we find might be underestimates of the level of bias present in more serious cases. While a valid concern, we believe that our results remain interesting for the following reasons. First, many of the previous studies on bias and discrimination in jury trials have focused on capital crimes. However, we believe that it is important to assess potential impediments to fairness across various types of trials/crimes that are much more prevalent and frequent in courtrooms. Second, in all of our regression specifications, we control for the type of crime. Moreover, in alternative specifications, we interact our main variables of interest with an indicator of whether the crime is a murder or a sex crime and find no significant differences in our main results across case types.

Panel B shows the case-level means of the various sets of controls on which we rely in our empirical analysis. The case descriptive variables show that the majority of the cases have black defendants and black victims. Over half of the cases are represented by a public defender, which generally speaks to the low income levels of the defendants. There is wide variation in the voir dire process, from who conducted the voir dire[22] to the length of time it took to shape the final jury. A Batson objection was raised in about 14% of the cases. A typical case in our sample involves close to three different charges against the defendant(s). Of these charges, about 56% result in a conviction, 38% in an acquittal, and the remaining 6% in a hung jury.

The next set of variables summarizes the average characteristics of the seated jury. These are our main variables of interest. In the jury supplemental survey, jurors only provide categorical responses to questions about their education, income, and age. In order to facilitate a more useful interpretation of our coefficients, we translate the categorial variables into appropriate levels for each variable by taking the mid-point of each category. However, using the categorical variable directly does not change our main results. In our restricted sample, a typical seated jury has jurors with about 15 years of education who earn $58,000 in annual income and are about 42 years old. On average, roughly 66% of the jurors seated for

[22]"Who conducted" is a variable ranging from 1 to 4, increasing in attorney involvement, where 1 is "Judge with little or no attorney involvement" and 4 is "Attorney with little or no judge involvement."

Table 1: Case Summary Statistics

Panel A:

Case Type	% of Cases	N
Murder (1st, 2nd, attempted) and Manslaughter	16.18	22
Sexual crime	2.94	4
Robbery, Burglary, Larceny, Theft, Assault, Arson	39.71	54
Others, including drug-related crimes	41.18	56

Panel B:

Variable	Mean	Std. Dev.	N
Case Characteristics:			
Victim	0.594	0.493	138
Female Victim	0.231	0.423	134
Non-black Victim	0.333	0.473	135
Female Defendant	0.098	0.299	132
Black Defendant	0.555	0.499	137
Public Defense	0.519	0.502	133
Total Number of Charges	2.713	2.277	136
Number of Charges Convicted	1.415	1.975	135
Voir Dire Characteristics:			
Who Conducted (1 "Judge" to 4 "Attorney")	2.309	0.923	136
Questionnaire Used	0.223	0.418	139
Voir Dire Time (hours)	5.343	6.278	137
Anonymous Jury	0.273	0.447	139
Batson Objection Raised	0.137	0.345	139
Jury Characteristics:			
Education (years)	15.101	1.090	139
Income ($, thousands)	57.603	12.462	139
Age	41.523	3.982	139
Religious	0.663	0.160	139
Female	0.569	0.179	139
Black	0.261	0.235	139
Woman Foreperson	0.508	0.502	130
Black Foreperson	0.248	0.434	133
Attorney Characteristics:			
Defense:			
Voir Dire Satisfactory (1 "Low" to 7 "High")	5.403	1.483	139
Legal Practice (years)	12.903	8.459	134
Previous Criminal Trials	62.323	73.089	133
Age	41.880	10.239	133
Female	0.313	0.465	131
Prosecution:			
Voir Dire Satisfactory	5.396	1.377	111
Legal Practice (years)	8.696	6.569	112
Previous Criminal Trials	39.584	57.515	113
Age	36.372	7.913	113
Female	0.495	0.502	105

Notes: Sample limited to cases with more than 5 jurors responding to school, income, age, race, gender, and religion questions, and non-missing site information. Also restricted to cases in which both the defense and the prosecution answered voir dire satisfaction question before learning verdict.

a particular case describe themselves as religious.[23] A typical jury in our restricted sample is over 50% women and about a quarter black, and the sex and race of the foreperson generally reflect the proportions in the overall jury population.[24]

Throughout our empirical analysis, we estimate specifications in which we control for attorney characteristics, because the attorney's satisfaction with the voir dire is a self-reported variable and his/her experience or sex may influence his/her evaluation of the seated jury. The mean level of satisfaction is the same for the defense as for the prosecution, and the dispersion of the satisfaction measure is similar for the two sides as well.[25] A typical defense attorney in our sample is older, has much more legal experience, and is more likely to be male compared to a typical prosecutor. Unfortunately, our data do not include the race of the attorneys.

4 Juror Characteristics and Biases

In this section, we first explore the link between individual juror characteristics and the pre-deliberation opinions held by jurors. Next, we examine the role of jury composition in shaping trial outcomes by assessing the relationship between average jury characteristics and verdicts. We then put these jury composition effects on verdicts in the context of the effect of evidentiary strength, treating the judge's assessment of the trial evidence as an objective evaluation of the case against the defendant.

4.1 Pre-Deliberative Juror Biases

The survey of jurors in our dataset asks a series of questions about opinions concerning the defendant, the victim, the attorneys, and other aspects of the case, as well as crime more generally. We ask whether differences in these stated opinions across jurors within each case can be explained by the jurors' demographic and socioeconomic characteristics. More

[23] Answered 1 or 2 on a scale of 1 to 5 where 1 is very religious, 2 is religious, and 5 is very nonreligious.

[24] Comparing our restricted sample with the county population demographics, we find that females are overrepresented in our sample in Bronx (62.1% in NCSC sample versus 53.4% in the Census) and D.C. (64.3% in the NCSC sample versus 52.9% in the Census) while blacks are overrepresented in L.A. (15.8% versus 9.4%) and Bronx (40.4% versus 31.9%) and underrepresented in D.C. (44.9% versus 59.1%) The median household income of the seated jurors in our sample are significantly higher than in the county population estimates from the Census by about $15,000 to $20,000. These differences can be partially attributed to the selective pools from which initial jury summonses are drawn. County demographic estimates are from U.S. Census Bureau Population Estimates, April 1, 2000 to July 1, 2010. Retrieved March 8, 2011 from `http://www.census.gov/popest/estimates.html`. Household median income data from U.S. Census Bureau Small Area Income and Poverty Estimates. Retrieved March 15, 2011 from `http://www.census.gov/did/www/saipe/data/index.html`.

[25] We will discuss this variable in more detail in Section 5.

specifically, we estimate the case fixed effects model

$$y_{jc} = \beta X_{jc} + \alpha_c + \varepsilon_{jc}, \tag{1}$$

where y_{jc} is the level of agreement of juror j with various statements about case c and the parties involved (on a scale ranging from 1 "Strongly Disagree" to 7 "Strongly Agree"), X_{jc} is a vector of juror characteristics and selected interactions with some case characteristics, and α_c captures case-level characteristics.[26] Introducing case fixed effects ensures that variations in jurors' opinions cannot be attributed to differences in unobservable case-level characteristics.

Table 2 reports OLS estimates of equation (1) for six separate questions that jurors were asked to state an opinion on. While this linear modeling approach ignores the ordinal nature of the dependent variables, alternative models that would account for this ordinality, such as the ordered logit or ordered probit, encounter the incidental parameters problem within a fixed effects framework, which renders the usual maximum likelihood estimators inconsistent even when the models are properly specified (Greene and Hensher, 2010). We have chosen to focus on the results from the linear fixed effects models to ensure that our estimates are reliable and free from the confounding effects of unobserved within-case factors.[27]

Table 2 provides evidence of significant relations between the education, income, and religiousness of individual jurors – characteristics that have been largely ignored in past studies on jury discrimination – and their pre-deliberation leanings. Compared to others who heard the same case, jurors with higher education levels are less likely to express sympathy for the victim (column 1). More educated jurors are also significantly more likely to interpret the trial evidence as strongly favoring the prosecution relative to the defense (column 6). Specifically, an extra year of education is associated with a 0.03 point decrease on the 7-point scale that measures juror interpretation of evidentiary strength.[28]

Most consistently across the six predispositions, jurors with higher income hold pre-deliberation leanings that are more favorable for the prosecution. While individuals with

[26] We cannot observe whether the same jurors appear in more than one case. However, given the short period of data collection in each county, it is highly unlikely that one individual would have been summonsed for jury duty more than once within that period, let alone making it on to the seated jury more than once. We therefore treat each observation in our dataset as representing a distinct juror.

[27] A solution that allows both unobserved heterogeneity and ordinality to be accounted for is presented by Baetschmann et al. (2011), who propose and demonstrate the consistency and favorable small-sample properties of an alternative fixed effects ordered logit estimator. Utilizing this BUC estimator produces estimates that are qualitatively similar though less precise than those from our linear models. Note that each specification in Table 2 is estimated independently: attempts to account for cross-equation correlations also produced less precise but substantively similar estimates.

[28] The dependent variable in column 6 is the juror's response to the question "All things considered, how close was the case?" to which the juror can answer on a scale of 1 to 7 where 1 is "Evidence strongly favored the prosecution" and 7 is "Evidence strongly favored the defense."

higher earnings are more critical of attorney skills for both the prosecution and the defense, we find that higher income is significantly associated with less sympathy for the victim (column 1), greater trust in police in the community (column 2), and a greater likelihood of evaluating the trial evidence in favor of the prosecution (column 6). Specifically, a $10,000 increase in juror income has the same effect as an additional year of education (i.e. a 0.03 point decrease on the 7-point scale) on the interpretation of the trial evidence. Similarly, jurors who self-identify as "Religious" or "Very Religious" express significantly greater trust in the police, a higher rating of the prosecution's case (column 4), and an interpretation of the evidence inclined towards the prosecution. Thus, higher education, income, and religiousness appear to be associated with a general bias towards the prosecution.

Table 2 also reveals important sex and race effects on individual juror biases before group discussion. For clarity and ease of interpretation, the middle panel in Table 2 reports the full sex and race effects under different sex/race composition identities of the victim and defendants in the case.[29] Female jurors tend to rate the skill of the prosecuting attorney more highly (column 3) in all cases, but appear to have some other biases that depend to some degree on the gender of the victim. For example, female jurors are weakly less likely than other jurors to report sympathy for the victim or a high rating of the defense's overall case when the victim is male, but display a smaller difference in these regards compared to other jurors when the victim is female. However, the coefficents are imprecisely estimated, and equality of the gender effects by sex of the victim cannot be ruled out. While firm conclusions are more difficult to draw for the case of juror gender, the greater tendency to rate the prosecutor's skill highly and to regard the defense's case as weak indicate that women may have a moderate general bias towards the prosecution.

Examining the relation between juror race and pre-deliberative biases, we find large effects consistently across the six predispositions and most combinations of defendant and victim race. Black jurors report significantly lower trust in the police, lower ratings of the prosecutor's skill and case, higher ratings of the defense's case, and an interpretation of the evidence more in favor of the defense. The divergence in evaluations of the prosecutor's case and the defense's case by black jurors relative to other jurors is most prominent when the defendant is black. On the other hand, the divergence in the evaluation of the prosecutor's skill by black jurors relative to other jurors is most prominent when the defendant is *not* black. Finally, although the race effect on the interpretation of the evidence is only statistically significant when both the defendant and the victim are black, the magnitudes are similar for cases involving all defendant/victim race combinations. Specifically, black

[29]For example, the full black effect when the defendant and the victim are black (i.e. dblack = 1 and vnblack = 0) is the sum of coefficients on black and black × dblack.

Table 2: Pre-Deliberation Juror Opinions with Case Fixed Effects: OLS Results

	Dep. Var. = 1 to 7, where 1 is "Strongly Disagree" and 7 is "Strongly Agree"					
	Sympathy for Victim	Trust Police	Skillful Prosecutor	Prosecution Case Strong	Defense Case Strong	How Close (P → D)
	(1)	(2)	(3)	(4)	(5)	(6)
education (years)	-0.113**	-0.001	0.008	0.016	-0.008	-0.031*
	(0.053)	(0.018)	(0.020)	(0.016)	(0.017)	(0.016)
income ($, thousands)	-0.013***	0.005***	-0.004**	-0.000	-0.004***	-0.003*
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
age	-0.000	0.002	0.003	-0.001	-0.003	-0.002
	(0.010)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
religious	0.027	0.251***	0.050	0.211***	0.027	-0.148*
	(0.225)	(0.084)	(0.090)	(0.081)	(0.081)	(0.082)
female	-0.560	0.076	0.307**	0.066	-0.319*	-0.089
	(0.443)	(0.153)	(0.133)	(0.143)	(0.162)	(0.162)
female × vfemale	0.626	-0.064	0.063	-0.218	0.225	0.150
	(0.405)	(0.198)	(0.175)	(0.202)	(0.165)	(0.183)
black	0.593	-0.434*	-0.480**	-0.079	0.032	0.380
	(0.655)	(0.256)	(0.188)	(0.207)	(0.236)	(0.234)
black × dblack	-0.155	0.051	0.342	-0.243	0.194	-0.029
	(0.680)	(0.276)	(0.221)	(0.226)	(0.254)	(0.261)
black × vnblack	-0.272	-0.336	0.051	-0.126	0.350	-0.109
	(0.666)	(0.300)	(0.245)	(0.235)	(0.290)	(0.286)
Female effect when						
vfemale = 0	-0.560	0.076	0.307**	0.066	-0.319*	-0.089
	(0.443)	(0.153)	(0.133)	(0.143)	(0.162)	(0.162)
vfemale = 1	0.066	0.012	0.371**	-0.153	-0.094	0.062
	(0.404)	(0.204)	(0.190)	(0.185)	(0.154)	(0.177)
Black effect when						
dblack = 0, vnblack = 0	0.593	-0.434*	-0.480**	-0.079	0.032	0.380
	(0.655)	(0.256)	(0.188)	(0.207)	(0.236)	(0.234)
dblack = 1, vnblack = 0	0.438	-0.383**	-0.138	-0.322**	0.227*	0.351**
	(0.379)	(0.162)	(0.160)	(0.141)	(0.132)	(0.147)
dblack = 0, vnblack = 1	0.320	-0.770***	-0.429*	-0.206	0.382*	0.271
	(0.555)	(0.212)	(0.237)	(0.225)	(0.236)	(0.233)
dblack = 1, vnblack = 1	0.166	-0.719**	-0.087	-0.449**	0.577*	0.241
	(0.664)	(0.312)	(0.254)	(2.227)	(0.295)	(0.292)
N	1418	1406	1400	1409	1403	1411
# of Cases	169	169	169	169	169	169
F-stat	1.693	3.603	1.296	1.245	2.420	2.102

Notes: Robust standard errors clustered at the case level are reported in parentheses. Sample limited to those cases with more than 5 jurors responding to school, income, age, race, gender, and and religious questions and to those cases in which both the defense attorney and the prosecutor answered the voir dire satisfaction question before learning verdict. "How close" asks "All things considered, how close was the case?" and the juror can answer from 1 to 7 where 1 is "Evidence strongly favored prosecution" and 7 "Evidence strongly favored defense." All regressions also control for the juror's employment status, previous jury experience, previous criminal trial experience, and his/her race and sex interacted with the race and sex of the foreperson. ∗=10%, ∗∗=5%, ∗ ∗ ∗=1%.

jurors report an interpretation of the evidence that is rougly 0.25 to 0.35 points higher on the 7-point scale than other jurors. To put the magnitude of this race effect in context, an equivalent change in the interpretation of evidence would require a drop in income of about $100,000 or a decline in education of about 10 years. Thus, black jurors appear to have a strong general bias towards the defense.

4.2 Average Jury Characteristics and Verdicts

Having demonstrated that there are significant relations between individual juror characteristics and pre-deliberation leanings, we now explore whether these individual biases feed through into biases in the determinations of guilt that juries make collectively. More precisely, we will assess the relation between the average characteristics of the seated jury and the trial verdicts. Before proceeding with the analysis, we first address some methodological issues.

While our dataset includes information on how individual jurors voted at various stages of the deliberations, we have chosen to focus on actual trial verdicts rather than individual juror voting patterns. This choice was motivated by several considerations. First, as noted by some legal scholars, the human dynamics involved in jury deliberation provide a potential channel through which individual prejudices can be attenuated or amplified within a group-decision setting.[30] We therefore consider the question of how jury composition relates to the verdicts arrived at when deliberations have concluded to be of much more policy relevance than the question of how individual juror characteristics relate to individual juror votes at earlier stages of deliberation or to how individual jurors would have decided the case on their own. Second, while estimating a case fixed effects model relating the voting patterns of individual jurors to their characteristics might mitigate bias due to unobserved heterogeneity, it would necessarily entail dropping all cases in which all jurors voted the same way. Since hung juries occur for only 8 percent of the counts in our estimation sample, this restriction would drastically reduce our sample size and statistical power. Finally, our dataset only contains information on individual jurors' votes on the "most serious charge" against the defendant. The wording of the survey questions appears to have led to differing interpretations across respondents and, more generally, a low response rate, which limits our confidence in the quality of this information.

[30]This idea is related to the notion of "group polarization" found in the psychology literature, which states that, if a group is like-minded, discussion strengthens its prevailing opinions. In an influential study, Myers and Bishop (1970) found that talking about racial issues increased prejudice in a high-prejudice group of high school students and decreased it in a low-prejudice group. Bringing this idea into the legal field, Sunstein (2000) writes that, in small deliberative groups such as juries, there may be a "robust pattern" of polarization whereby the initial inclinations of individuals before deliberation become more severe during deliberation.

Having thus decided to focus on verdicts rather than voting patterns, we must still confront the issue of how to precisely define and specify our dependent variable. Past studies – most notably Anwar et al. (2012a,b) – have focused on a binary dependent variable indicating whether there was a conviction on any count for a given trial. However, these studies report that few of the trials being examined had multiple charges. In contrast, Table 1 shows that the average number of counts per trial in our dataset is 2.7, with a wide variance around this mean. We consider the loss of information that would result from the collapse of the outcomes on such a large number of charges into a binary conviction indicator to be undesirable. The alternative that we have adopted is to represent trial outcomes with a pair of count variables, indicating the absolute number of convictions and acquittals for a given trial respectively. While we do not attempt to account for varying seriousness of charges or to otherwise weight individual counts, our approach allows us to distinguish outcomes more finely – and perhaps also more in line with the objectives of attorneys on each side – than an approach based on a binary indicator.[31] We model trial outcomes thus defined by way of an independent poisson regression for each dependent variable.[32]

Finally, we must return to the endogeneity concerns raised by the discussion of the voir dire process in Section 2. The unit of observation for our analysis of the impact of jury composition on trial outcomes will necessarily be individual cases, so that it will no longer be possible to employ case fixed effects.[33] This raises the possibility that our results will be

[31]Of the 2.7 counts per trial, 1.4 result in a conviction on average, while 1.1 result in an acquittal, and the remaining 0.2 result in a hung jury. In contrast, headline conviction rates for American criminal trials in general are typically reported as two thirds or higher (see, for example, the *Sourcebook of Criminal Justice Statistics Online*, Table 5.57). However, such headline conviction rates refer to convictions on any charge for a given trial, including secondary counts and pleas to lesser charges. The relatively low conviction rate suggested by the average number of acquittals and hung juries per count in our data is precisely the result of our decision to examine counts within a case individually rather than compressing this information into a simple any conviction/no conviction assessment for each trial.

[32]Our modeling choice has primarily been driven by efficiency concerns. The results that we present below remain qualitatively similar if we instead use a binary dependent variable, either estimated by way of a linear probability model as by Anwar et al. (2012b) or a logistic model; and are likewise very similar to those that are obtained from a linear or negative binomial model applied to the count dependent variables. However, the results from the poisson model are estimated the most precisely, and are therefore the results that we focus on. This decision also receives some support from various specification tests, in which the null hypothesis that the poisson model is the correct one cannot be rejected in our data. While the negative binomial model would allow us to relax the assumption implicit in the poisson model that the mean and variance of the dependent variable are the same (which, in any case, is not obviously false in our data), we instead address this potential concern by using robust standard errors within the poisson model, which has the additional attractive feature of being robust to misspecification. Unfortunately, our relatively small sample size also raises precision concerns when attempting to account for cross-equation correlation in our estimation, for example by way of the seemingly-unrelated poisson approach suggested by King (1989).

[33]Our discussion above on the number of charges per trial suggests that we could alternatively treat individual counts as the unit of observation, which would permit us to control for case fixed effects. However, since all of our explanatory variables – including, most notably, average jury characteristics – vary at the trial level only, this approach would not allow us to address our hypotheses of interest.

biased by unobservable factors related to both verdicts and jury composition. For example, attorneys might choose jury selection strategies based on observable juror characteristics jointly with strategies for other aspects of the trial, thereby inducing some simultaneity in the determination of jury composition and verdicts. We address this endogeneity issue by controlling for factors affecting attorney strategy and as many other aspects of the trial as possible. Our dataset contains a rich set of such proxy variables for many potential sources of unobserved heterogeneity.

Specifically, we control for all of the case, voir dire, and attorney characteristics that are summarized in Table 1.[34] The attorney experience variables, in conjunction with the voir dire characteristics, are meant to capture both the means and the ability of attorneys to design and implement jury selection strategies that could affect the composition of the seated jury. The remaining case-level variables are meant to capture both the direct effects of these and other characteristics with which they are correlated, as well as any indirect effects that may result from attorneys attempting to condition their strategies on these variables.

In addition, we include a set of variables describing the trial judge's opinion of various aspects of the case, including strength of evidence, attorney performance for each side, and the complexity of the evidence and the law. Table 3 summarizes the seven variables accounting for these judge opinions, which we treat as objective assessments of these aspects of the trials.[35] The attorney performance variables provide additional indicators of attorney ability to shape jury composition and trial outcomes, while the other aspects, in addition to their direct interest, may also guide or constrain attorney strategies. Moreover, the evidence variables provide us with a unique opportunity to assess the relative prominence of jury composition effects as drivers of trial outcomes.[36]

Table 4 reports our poisson results for the number of convictions and the number of acquittals. To avoid over-fitting our data and to preserve as many observations as possible, we include some specifications that limit the control variables included. Columns 1 and 4 and columns 2 and 5 control for measures of the presiding judge's opinion of the case and attorney characteristics, respectively, while columns 3 and 6 include both sets of judge and attorney controls. Moving from left to right, the number of trials for which data on all

[34]The voir dire satisfaction variables are excluded. We treat voir dire satisfaction as a separate outcome variable, and analyze its determinants in detail in the following section.

[35]We do not observe identifying or demographic information about the trial judges. It is therefore possible, in the absence of other proxies for judge effects, that these judge assessments – as well as the voir dire characteristics, since, as discussed above, judges have a high degree of discretion over the voir dire process – will also capture the effects of otherwise unobservable judge characteristics. Given the size of the jurisdictions covered by our dataset, we regard the number of trials presided over by a single judge to be small, though we ultimately have no way to verify this.

[36]Though measured on a seven-point Likert scale, we include these variables linearly in our main specifications. Alternative categorical specifications do not alter our results.

Table 3: Summary of Judge Opinions

Variable	Mean	Std. Dev.	N
Evidence presented at trial complex (1 = "Not at all" to 7 = "Very complex")	2.328	1.526	134
How complex was the law (1 = "Not at all" to 7 = "Very complex")	2.597	1.576	134
Evidence favored which side (1 = "Prosecution strongly" to 7 = "Defense strongly")	3.120	1.360	133
Attorneys presented all relevant evidence (1 = "Completely disagree" to 7 = "Completely agree")	5.561	1.540	132
Skillful prosecutor during trial (1 = "Not at all" to 7 = "Very skillful")	5.000	1.535	136
Skillful defense attorney during trial (1 = "Not at all" to 7 = "Very skillful")	5.199	1.444	136
How important was police testimony (1 = "Not at all" to 7 = "Very important")	5.052	1.669	134

Notes: Sample limited to cases with more than 5 jurors responding to school, income, age, race, gender, and religious questions. Sample also limited to those cases in which both attorneys answered voir dire satisfaction question before learning verdict.

control variables in a given specification are available decreases. Restricting the sample to only those cases included in the specifications in columns 3 and 6 does not change the results in the other columns.[37]

Our results reveal that some of the pre-deliberation biases we found to be associated with various juror characteristics do indeed appear to feed through into the verdicts that juries with concentrations of these juror characteristics arrive at. Juries made up of higher income individuals are strongly associated with a decreased number of acquittals. In the specification controlling for both attorney characteristics and judge opinions (column 6), a $10,000 increase in the jury's average income is approximately associated with a 64 percent decrease in the

[37]Many control variables are suppressed in Table 4 to save space, but are included in the regressions. In addition to the variables already discussed, all equations also include controls for county, case type, and number of counts. Constraining the coefficient on the natural log of the total number of counts in a given trial to be one would impose the restriction that the rate of conviction or acquittal stays constant as the number of counts rises. Imposing this restriction makes little difference to the rest of our results, though the null hypothesis that it holds can be rejected at conventional significance levels. Specifically, the unconstrained coefficient estimate on the natural log of counts is greater than one in the acquittals equations and, correspondingly, less than one in the convictions equations, indicating that the rate of acquittal is actually increasing in the total number of counts. This may indicate that prosecutors add less serious charges when the main charge lacks evidence in an attempt to secure at least one conviction, or that defense attorneys focus their efforts on a large number of secondary charges when conviction on the main charge seems assured; either would lend further support to our assertion that a binary conviction indicator would seem to miss some potentially important information.

Table 4: The Effect of Average Jury Characteristics on Verdict: Poisson Results

	Dep. Var. = Number of counts that resulted in					
	Convictions			Acquittals		
	(1)	(2)	(3)	(4)	(5)	(6)
education (years)	-0.187*	-0.172	-0.085	0.211	0.204	0.000
	(0.100)	(0.158)	(0.173)	(0.129)	(0.204)	(0.169)
income ($, thousands)	0.010	0.002	0.002	-0.014	-0.035*	-0.064***
	(0.009)	(0.015)	(0.014)	(0.012)	(0.019)	(0.019)
age	-0.021	-0.064*	-0.091***	0.039	0.048	-0.005
	(0.017)	(0.034)	(0.030)	(0.025)	(0.040)	(0.026)
religious	0.446	1.036*	3.218**	-0.401	-1.990**	-1.893*
	(0.447)	(0.616)	(1.328)	(0.640)	(0.888)	(0.986)
female	-0.241	-0.646	-1.036	0.057	0.864	4.333***
	(0.624)	(0.806)	(1.002)	(0.713)	(0.816)	(1.067)
female × vfemale	-0.979	1.003	1.908	0.409	-1.174	2.278
	(1.205)	(1.741)	(3.068)	(1.310)	(3.196)	(3.295)
black	-1.029	-0.263	-2.370	2.294*	2.933*	7.163***
	(1.079)	(1.595)	(2.206)	(1.206)	(1.591)	(1.769)
black × dblack	-0.330	-3.293**	-2.000	-2.241**	-1.422	-6.210***
	(0.912)	(1.360)	(1.512)	(0.967)	(1.336)	(1.533)
black × vnblack	1.507*	1.814	3.859***	-1.367	-1.820	-7.544***
	(0.798)	(1.307)	(1.314)	(1.106)	(1.618)	(1.584)
Judge - Side Evidence Favored	-0.297***		-0.442***	0.374***		1.082***
(1 "Pros." to 7 "Def.")	(0.065)		(0.100)	(0.077)		(0.149)
Female effect when						
vfemale = 0	-0.241	-0.646	-1.036	0.057	0.864	4.333***
	(0.624)	(0.806)	(1.002)	(0.713)	(0.816)	(1.067)
vfemale = 1	-1.221	0.357	0.872	0.466	-0.311	6.611**
	(0.981)	(1.435)	(2.745)	(1.234)	(3.148)	(3.202)
Black effect when						
dblack = 0, vnblack = 0	-1.029	-0.263	-2.370	2.294*	2.933*	7.163***
	(1.079)	(1.595)	(2.206)	(1.206)	(1.591)	(1.769)
dblack = 1, vnblack = 0	-1.359*	-3.556***	-4.369***	0.052	1.511	0.953
	(0.708)	(1.114)	(1.273)	(0.841)	(1.416)	(0.888)
dblack = 0, vnblack = 1	0.477	1.551	1.489	0.927	1.113	-0.382
	(1.079)	(1.065)	(2.086)	(1.087)	(1.486)	(1.522)
dblack = 1, vnblack = 1	0.148	-1.742	-0.511	-1.314	-0.309	-6.591***
	(0.935)	(1.341)	(1.499)	(1.153)	(1.482)	(1.614)
N	165	98	93	165	98	93
Log-Likelihood	-202.969	-120.432	-99.444	-185.635	-92.642	-71.386
Judge's Opinion?	✓		✓	✓		✓
Attorney Controls?		✓	✓		✓	✓

Notes: Robust standard errors are reported in parentheses. Sample limited to cases with more than 5 jurors responding to school, income, age, race, gender, and religious questions. ∗=10%, ∗∗=5%, ∗ ∗ ∗=1%. All columns also include number of counts; case, voir dire, and foreperson characteristics; and county and case type fixed effects.

expected number of acquittals.[38] Likewise, a higher proportion of religious individuals on the jury is associated with a large increase in the expected number of convictions, and a large decrease in the expected number of acquittals, with a 10-percentage point increase in the proportion of religious jurors approximately associated with a 19 percent decrease in the expected number of acquittals.[39] These results directly correspond with the generally pro-prosecution bias we identified previously with religious and high-income individual jurors.

In contrast, some of the effects of jury composition on verdicts that we uncover do not correspond as directly with relations between the associated individual juror characteristics and leanings. Whereas jurors with more years of education are more likely to interpret evidence in favor of the prosecution, juries with a higher average level of education convict at a weakly lower rate and acquit at a weakly higher rate, though the coefficients are not estimated precisely. On the other hand, whereas juror age was neither statistically nor substantially associated with any individual juror leanings, the results in Table 4 indicate that juries with a higher average age convict on significantly fewer counts. This latter finding is notable because it stands in contrast to the findings of Anwar et al. (2012a), which suggest that older jurors are *more* likely to convict than younger jurors. One possible explanation for these differing conclusions about the effect of age is that Anwar et al. (2012a) cannot control for juror characteristics beyond age, race, and gender. Since older individuals are more likely to earn higher incomes, the age effect that Anwar et al. (2012a) find may be biased by the negative impact of income on acquittals, as they are forced to omit income from their analysis due to data limitations.[40]

As with our analysis of individual juror inclinations, it is difficult to draw firm conclusions concerning effects of the gender composition of juries on trial outcomes. There is some evidence that juries with a greater proportion of women are more lenient towards defendants

[38]Poisson coefficients are semi-elasticities, i.e. proportional changes in the dependent variable for an infinitessimal level change in the independent variable of interest. Of course, these are only approximate when the contemplated change in the independent variable is large. Poisson coefficients can alternatively be exponentiated to calculate what are known as the associated incidence-rate ratios, which give the exact ratio of the level of the dependent variable corresponding to a unit increase in the independent variable of interest to the level of the dependent variable corresponding to some baseline set of values of the regressors. We discuss our results in terms of semi-elasticities in order to retain a transparent link to the estimated coefficients reported in Table 4

[39]Increasing the proportion of the jury with a given characteristic by ten percentage points (i.e. by 0.1) is approximately equivalent to adding one person with that characteristic to a twelve-person jury.

[40]It should be mentioned, however, that the research design of Anwar et al. (2012a) allows them to isolate the effect of random variation in the age composition of juries on conviction rates. So while they are able to deal with endogeneity in return for potentially suffering from a separate omitted variable issue, our ability to include additional juror characteristics comes with a potential endogeneity issue that our proxy variable approach may not fully address. Our analysis of individual juror biases, which includes case fixed effects in addition to a full suite of juror characteristics, does not produce such directly contrasting results to the age effect identified by Anwar et al. (2012a), though neither does it provide corroborating evidence.

in terms of convicting on fewer counts and acquitting on more counts. This in contrast to the weak general bias in favor of the prosecution exhibited by individual female jurors that we discussed above. However, the effect on trial outcomes is only statistically significant in one specification. There are also signs that such leniency towards defendants is attenuated when the victim is female, but again, the effects are weak and imprecisely estimated.

Finally, we find racial composition effects on trial outcomes to depend primarily on victim race. In cases with a black victim, we find that a greater proportion of blacks on the jury is associated with significant increases in the number of acquittals and decreases in the number of convictions. The positive effect on the number of acquittals is most prominent when the defendant is not black, while the negative impact on convictions is strongest when the defendant is black. The effects are quite large, with a 10-percentage point increase in the proportion of blacks associated with a decrease of between 10 and 44 percent in convictions depending on the specification and the race of the defendant.[41] These results are consistent with the general pro-defense biases we noted amongst individual black jurors. When the victim is not black, the effects are imprecisely estimated, and vary widely in sign and magnitude across specifications.[42]

Although we find that many aspects of jury makeup are significant predictors of trial outcomes, we find that more substantive aspects of the trial are also strong predictors of verdicts. For example, the judge's opinion regarding the evidence presented at trial as favoring the defense versus the prosecution has a large impact on the number of convictions and acquittals, in the expected direction.[43]

Even though we cannot know what the "true" or "correct" outcome should have been for any of the trials in our dataset, we can use this measure of the strength of evidence to put the magnitude of the jury composition effects in the context of factors that are closer to core notions of justice. Suppose that there could never be evidence stronger than one

[41]Note that, whereas the variable 'black' referred to an indicator of the individual juror's race in Table 2, it refers to the proportion of the jury that is black in Table 4.

[42]Perhaps surprisingly, in a case with a black defendant and a non-black victim, our results from the specification including all controls suggest that a greater proportion of blacks on the jury is associated with a substantial and significant *decrease* in the number of acquittals. One possible explanation for this somewhat surprising result may be that the victim not black indicator has a composite excluded category capturing all cases with a black victim as well as cases without a victim, and that black jurors are only more likely to acquit in cases without a victim. We do, however, control directly for whether there is a victim in the case, which mitigates this concern; and specifications with alternative victim variable definitions and interactions do not produce qualitatively different findings. A more plausible explanation is that the specification in column 6 is over-fitting the data, which is supported by the small number of included trials with non-black defendants and/or non-black victims and the corresponding large jumps in the magnitude and significance of the race coefficients moving to this columng from the other specifications. We therefore do not regard this particular estimate as reliable or indicative of a general effect.

[43]While not shown in Table 4 in consideration of space, we also find that the judge's assessments of other characteristics of the trial and evidence are significant correlates of verdicts.

standard deviation below the mean strength of evidence against any defendant who is truly innocent, and, symmetrically, that there could never be evidence weaker than one standard deviation above the mean against a defendant who is truly guilty. Then, for a truly guilty defendant to be acquitted, factors besides the strength of evidence would have to have an effect equivalent to at least a two-standard-deviation shift in the strength of evidence. As shown in Table 3, the sample standard deviation of the judge's assessment of the strength of evidence is 1.36, so that the estimates in Table 4 imply that a two-standard-deviation shift in the strength of evidence in favor of the defendant would be associated with an 81-120 percent decrease in the number of convictions per trial and a 102-294 percent increase in the number of acquittals. To achieve the same effect through jury composition would require, other things equal, a decrease in the average income of the jury by between $16,000 and $46,000, or an increase in the number of blacks on the jury of between 3 and 6. Or equivalently, to make a defendant confronted with evidence of average strength appear truly innocent or truly guilty, the race of roughly one third of the seated twelve-person jury would have to be changed, or the average income of the jury would have to be changed by roughly 40% starting from the sample average jury income level.

An alternative way to place the jury composition effects in perspective is to employ the regression-based technique proposed by Fields (2004) that decomposes the explanatory power of of an overall model into the contributions of the constituent independent variables. The basic idea is as follows. Let $s(X^k)$ be the share of the variation in the dependent variable Y that can be attributed to variation in the k-th independent variable. Fields defines $s(X^k)$ as

$$s(X^k) \equiv \frac{cov(X^k \hat{\beta}^k, Y)}{var(Y)}, \tag{2}$$

where $\hat{\beta}^k$ is the estimated coefficient on X^k from our regression of choice. Then $\sum_{k=1}^{K} s(X^k) = R^2$ where K is the total number of independent variables in the model. We can then express the $s(X^k)$ in terms of their percentage contribution to the R^2 of the regression model:

$$p(X^k) \equiv \frac{s(X^k)}{R^2}. \tag{3}$$

Performing the Fields (2004) regression-based decomposition thus allows us to compare the relative explanatory power of jury composition with that of evidentiary strength.[44] For

[44]In the poisson model, the traditional R^2 is not well defined, so the decomposition must be performed using the pseudo-R^2. An alternative would be to perform the decomposition for the analogous linear model. This alternative is suggested by Fields (2004), though he states that the practical differences tend to be minor for count data applications. We have likewise found little qualitative difference in results when applying the

convictions, variation in jury composition – in terms of education, income, age, religiousness, sex, and race – accounts for about 39% of the explained variation in our model in column 3, about half of which is from the effect of race. However, roughly 43% of the explained variation in convictions can be attributed to the set of variables associated with the judge's opinion of the evidentiary complexity and strength. For acquittals, 33% of the explained variation can be attributed to jury composition, and the judge's evaluation of the case's legal and evidentiary complexity and strength account for about 44% of the explained variation. Therefore, while the proportion of variation in trial outcomes explained by jury composition is not trivial, it is nonetheless exceeded by the proportion of variation explained by the more fundamental aspects of fact and legal context.

4.3 Summary

We have demonstrated in this section that there is some basis for concerns that jury composition can affect whether justice is achieved in criminal trials. Opinions, sympathies, and inclinations that would lead a juror to be biased towards one side are indeed correlated with observable juror characteristics. And these individual biases do indeed appear to feed through into verdicts when juries are composed of a larger number of individual jurors with the associated observable characteristics.

However, it is important to note that the relevant observable characteristics are not restricted to race, which has been the primary focus of several past studies. Most notably, we also find income to be an important predictor of individual biases and determinant of trial outcomes. While this is a novel and potentially important finding on its own, it also raises the possibility that the race effects identified in past studies are confounded with effects that are properly attributed to separate characteristics excluded from the analysis. Further, these jury composition effects are smaller and have less explanatory power than effects associated with the more fundamental aspects of the trial, i.e. those related to fact and the relevant legal doctrine.

Nonetheless, concerns that jury composition effects – regardless of size or source – can impede fairness in criminal trials remain. Furthermore, they raise a potential further concern, which we now proceed to evaluate: whether attorneys are aware of these jury composition effects, and might therefore attempt to make use of them to manipulate trial outcomes.

decomposition to a linear model.

5 Attorney Preferences Over Juror Characteristics

Motivated by our findings that there exist significant and robust relations between juror characteristics – especially income, religiousness, and race – and both individual juror biases and verdict patterns, we now turn to assessing whether attorneys anticipate these relations. If attorneys are unaware of these effects, this may mitigate concerns of their impact on justice (and would certainly mitigate concerns of endogeneity in our trial outcomes specifications).

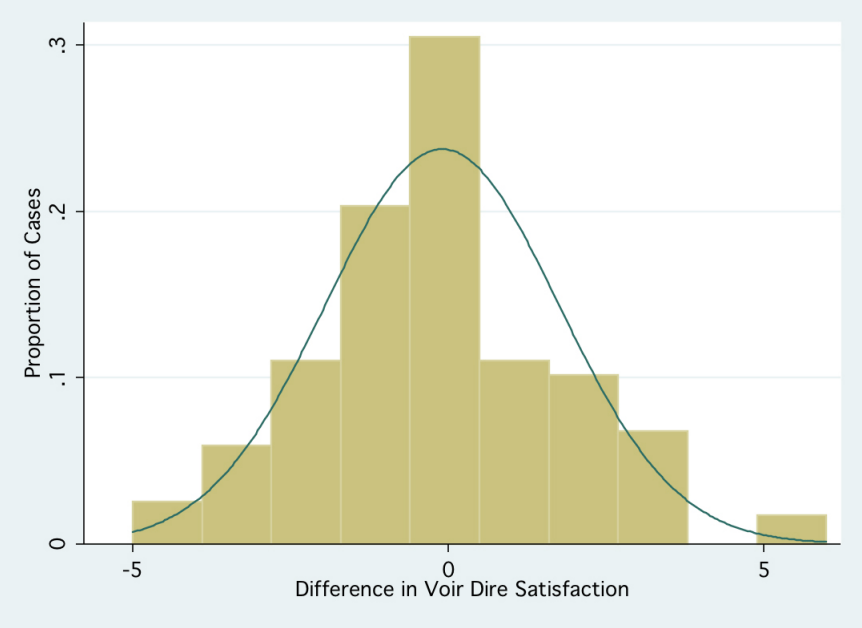
If, however, attorneys are aware of these effects, we would expect them to prefer seated juries with greater proportions of jurors exhibiting characteristics that we have found to be amenable to their side. In order to analyze attorney preferences over jury composition and assess if this is true, we capitalize on a self-reported measure of satisfaction concerning the jury selection process that attorneys were requested to state before learning the verdict. In the supplementary attorney survey, attorneys responded to the question, "How adequate was the voir dire in this trial?" on a scale of 1 to 7, where 1 is "Very inadequate" and 7 is "Very adequate." We interpret higher responses on this scale as indicating that the responding attorney felt better able to achieve a subjectively desirable jury composition.[45] We then ask whether this proxy measure for attorney satisfaction with the seated jury itself can be explained by variation in the average characteristics of the seated jury, controlling for case characteristics, various aspects of the voir dire process, attorney characteristics, and county fixed effects.[46]

Before proceeding with the analysis, we take a closer look at the voir dire satisfaction variables. Figure 1 illustrates the distribution across cases of the difference between the defense attorney's level of satisfaction and that of the prosecuting attorney. In about a third of the cases, the defense and the prosecution report the same level of satisfaction, and in another third of the cases, there is a one-point difference between the prosecution and the defense. While one might expect to see a greater proportion of cases with more polarized satisfaction, the shape of the distribution in Figure 1 could be driven by a few factors. First, one strategy that both attorneys might follow is to exercise their strikes against the

[45] In all of our analyses, we restrict our sample to trials for which the attorneys explicitly stated that they had responded to the voir dire satisfaction question before learning the verdict. This supports our interpretation of these satisfaction ratings by making it less likely that they simply capture post hoc rationalization of the broader trial outcome.

[46] It is important to note that we do not attempt to examine the degree to which attorneys actually manage to transform the composition of the jury in relation to that of the jury pool, nor do we attempt to identify the precise actions taken or strategies employed by attorneys in voir dire. While these questions are interesting in their own right, their answers depend both on attorney preferences and on a number of institutional and trial-specific constraints, and it is the former that we wish to focus on exclusively here. In Lehmann and Smith (2012), we develop a model of attorney behavior in voir dire to examine jury selection strategies and outcomes in detail.

Figure 1: Difference in Voir Dire Satisfaction within a Case: Defense Minus Prosecution



Notes: Sample limited to cases with more than 5 jurors responding to school, income, age, race, gender, and religious questions and those cases in which both defense and prosecution answer voir dire satisfaction question before learning the verdict.

potential jurors whom they perceive as having the most extreme prejudices against their side. After all strikes have been exercised, most of those who are left on the seated jury should then be jurors who are perceived as harboring only moderate inclinations to one side or the other, and are therefore not associated with an extreme preference held by either attorney. Second, if attorneys' jury composition preferences were simply based on a single characteristic of the jurors – for example, race – then one might expect a greater divergence in satisfaction with the seated jury, since one side's gain would necessarily be the other side's loss. However, if preferences are not unidimensional, then it is not clear that the prosecution and the defense should express such antithetical satisfaction with the overall makeup of the jury. For example, although the racial makeup of a given jury may benefit the prosecution, the religiousness and income levels of the jurors might be simultaneously more favorable to the defense. Finally, we do not know the benchmark against which any given attorney might express his or her satisfaction. So, for example, in a trial in which the jury pool has an unusually large proportion of blacks, the defense attorney might express a high level of satisfaction for having a greater proportion of blacks on the seated jury than in previous trials, while the prosecuting attorney might also express a high level of satisfaction

for successfully avoiding a seated jury with as many blacks as might have resulted from the particular jury pool.

The empirical approach we take is to estimate separate ordered logistic models for defense attorney satisfaction and prosecuting attorney satisfaction. This decision was based on a careful evaluation of the strengths and weaknesses of alternative approaches, and on various constraints imposed by our data.

The primary drawback of estimating separate models by attorney type is that it deprives us of the ability to leverage the limited panel structure of our data to address unobserved case-level heterogeneity. Since, in principle, we observe attorney satisfaction for each attorney type within a case, we should be able to treat our dataset as a panel, with trial as the cross-sectional dimension and two attorney-type observations for each cross-sectional unit. In practice, however, there are unfortunately only 80-90 cases in our dataset for which both attorneys responded to the voir dire satisfaction question before learning the verdict and for which various other control variables are also available. The panel is hence unbalanced to a substantial degree, though this in itself does not directly prevent us from employing a fixed effects or related strategy. Rather, our main reason for avoiding panel methods is that they render other important goals of our analysis difficult or impossible to achieve.

We would ideally like to account for the ordinal nature of the attorney satisfaction ratings, which the ordered logistic form accomplishes. However, the ordered logistic model with fixed effects leads to inconsistent estimates due to the incidental parameters problem. This can be overcome by instead employing the consistent fixed effects ordered logistic estimator proposed by Baetschmann et al. (2011), but the number of cases for which the associated conditional likelihood function exists is limited by the unbalanced nature of the panel and by the number of trials for which the prosecuting and defense attorneys reported the same values for the dependent satisfaction variable. Taking this approach hence reduces our sample size dramatically, leading to very imprecise estimates and susceptibility to over-fitting.

On the other hand, when we estimate separate linear specifications, implicitly treating the dependent variables as continuous cardinal measures of attorney satisfaction, the results do not differ substantially from the separate ordered logistic results that we present below. Since this suggests that it might be relatively costless to simply ignore the non-cardinality of the attorney satisfaction ratings, a seemingly natural way to proceed would be to pursue a linear fixed effects approach. Such an approach would not give rise to the incidental parameters problem, would not require a balanced panel – thus allowing us to retain as many observations as possible – and is indeed the very approach we pursued in our analysis of individual juror pre-deliberative biases. However, in the case of attorney satisfaction, a major drawback arises with *any* of these fixed effects methods: since the explanatory variables of interest (namely,

average jury characteristics) vary only at the trial level, it becomes impossible to identify the absolute effects of these variables on defense and prosecuting attorney satisfaction when controlling for case fixed effects. Of course, by interacting the average jury characteristics with a defense attorney indicator, we could estimate the effects of these characteristics on defense attorney satisfaction *relative* to prosecutor satisfaction; but in addition to such estimates having a less straightforward interpretation, they would not allow us to test our hypotheses of interest. Having found, for example, that juries with higher average incomes acquit on fewer counts, the natural prediction is that defense attorneys, if they are aware of this effect, should, other things equal, be absolutely less satisfied with a jury with a higher average income, not just relatively less satisfied than the prosecutor.

Our preferred choice in navigating these trade-offs is to focus on the separate ordered logistic models for each side, which permits us to account for the ordinal nature of the dependent variables, but more importantly, to identify the absolute effects of average jury characteristics on the satisfaction of each attorney type. Moreover, this approach also allows us to retain the greatest number of observations possible, thus leading to no disadvantage in terms of the precision of the estimates. Unfortunately, this method leaves open the possibility that our estimates will be biased by unobserved case-level factors that simultaneously affect attorney satisfaction and jury composition. As with our analysis of the effects of average jury characteristics on verdicts, we attempt to address such endogeneity concerns by controlling for a rich set of trial-level covariates. Specifically, besides the jury characteristics of primary interest and the characteristics of the attorney of the given side when applicable, each specification also includes our full set of case and voir dire characteristics. In addition, we have attempted to diagnose the presence of any remaining unobserved heterogeneity by estimating alternative panel models – despite the difficulties just mentioned that arise with such methods – and comparing the results to our baseline estimates. While the significance and apparent magnitude of some effects vary widely across these alternative specifications, the main results that we focus on below are relatively stable. We have therefore opted to report our baseline results from the separate ordered logistic specifications only.[47]

[47]The first set of alternative models combines the conditional binomial logit model with various methods to collapse the satisfaction ratings from a seven-point to a two-point scale. Estimation of these models relies on as few as 40 trials in some specifications. A somewhat more useful set of methods involves using the satisfaction of one side to attempt to capture case fixed effects in specifications explaining the other side's satisfaction. One way in which we implemented this idea was to treat the satisfaction ratings as cardinal and use their difference as the dependent variable. Alternatively, we included one side's satisfaction directly in the opposing side's specification, entering as a set of dummy variables or simply linearly. Finally, we implemented the linear fixed effects model with defense attorney interactions. We also made some attempts to increase efficiency, first of all by pooling observations across attorneys and constraining some coefficients – including, as in some other methods, the estimated logistic cut-off points on the unobserved latent satisfaction scale corresponding to the observed categorical variables – to be the same for both sides (and without including case

These results are presented in Table 5. The first three columns examine the determinants of the defense attorney's satisfaction, and the last three columns report the correlates of the prosecutor's satisfaction. Columns 1 and 4 control for all variables summarized in Table 1 except for attorney and foreperson characteristics, which are excluded in order to retain as many observations as possible. Specifications with these additional controls are reported in the remaining columns, and demonstrate the robustness of our main results to these different sets of controls.

The most striking finding in Table 5 is that the defense and prosecuting attorneys' levels of satisfaction are significantly correlated with the average income level of the seated jury, in opposite directions and in line with our findings on individual biases and trial outcomes. Seated juries with lower average income are associated with higher levels of satisfaction for the defense, while the prosecution's satisfaction level rises with a richer jury. This corresponds directly with our findings of a general pro-prosecution bias amongst individual jurors with higher incomes and the tendency for juries with higher average incomes to acquit on fewer counts. The significance and robustness of the estimates provide a strong indication that attorneys are well aware of these effects.

A second striking finding is that defense attorneys have a strong preference for seated juries with a high proportion of women. This preference is most prominent when there is no victim or a male victim, though the effect is still positive, if much less precisely estimated, when there is a female victim. This result is puzzling in relation to the effects of juror gender on individual biases and trial outcomes. While, as discussed above, juries with a greater proportion of women do appear to be weakly more lenient towards defendants, the effects are imprecisely estimated and somewhat sensitive to specification. Moreover, individual female jurors appear to have at least a weak general bias towards the prosecution. In contrast, the estimates in Table 5 indicate that defense attorneys have a very strong preference for female jurors.

One potential explanation for this apparent divergence between our previous results and attorney preferences is that attorneys use gender as an indicator of juror characteristics that they have trouble observing directly. For example, attorneys might expect that women tend to have lower incomes and fewer years of schooling or to be less religious, and – consistent with the results in Table 2 showing that these traits are associated with a greater tendency to interpret evidence in favor of the defense – prefer juries with a greater proportion of women

fixed effects); and, finally, by attempting to account for cross-attorney-equation correlation in a bivariate probit model, again with collapsed binary satisfaction ratings. Each of these specifications is adversely affected by the small sample sizes, whether through convergence problems, imprecise estimates, or nearly exactly determined coefficients for some explanatory variables. However, the two main results that we focus on are present and stable in sign and significance across all specifications.

Table 5: Attorney Voir Dire Satisfaction: Ordered Logit Results

Dep. Var. = Attorney Satisfaction with Voir Dire (1 to 7 where 1=Low and 7=High)						
	Defense			Prosecution		
	(1)	(2)	(3)	(4)	(5)	(6)
education (years)	0.239	0.213	0.152	-0.432*	-0.360	-0.344
	(0.204)	(0.218)	(0.242)	(0.252)	(0.260)	(0.299)
income ($, thousands)	-0.046**	-0.054***	-0.060**	0.067***	0.076***	0.078**
	(0.018)	(0.020)	(0.023)	(0.025)	(0.028)	(0.031)
age	-0.017	-0.018	-0.006	-0.001	-0.000	0.017
	(0.054)	(0.063)	(0.067)	(0.051)	(0.064)	(0.070)
religious	1.548	1.851	0.703	0.555	0.955	0.465
	(1.207)	(1.290)	(1.497)	(1.040)	(1.158)	(1.202)
female	3.678***	4.183***	4.684***	0.784	-1.427	-0.016
	(1.111)	(1.209)	(1.438)	(1.972)	(2.004)	(2.017)
female × vfemale	-0.203	-0.062	1.022	-3.376	-3.664	-7.122
	(3.480)	(3.138)	(3.499)	(3.195)	(3.777)	(4.818)
black	1.026	2.085	1.641	5.432*	6.096*	6.331*
	(2.457)	(2.708)	(3.120)	(3.254)	(3.161)	(3.720)
black × dblack	-2.634	-4.182	-4.401	-5.640**	-6.371**	-5.862*
	(2.480)	(2.780)	(2.906)	(2.857)	(2.770)	(3.232)
black × vnblack	0.883	0.368	1.353	-0.895	-1.227	-1.451
	(2.491)	(2.868)	(2.790)	(2.077)	(2.077)	(2.772)
Female effect when						
vfemale = 0	3.678***	4.183***	4.684***	0.784	-1.427	-0.016
	(1.111)	(1.209)	(1.438)	(1.972)	(2.004)	(2.017)
vfemale = 1	3.475	4.121	5.706*	-2.592	-5.090	-7.139*
	(3.268)	(2.892)	(3.351)	(2.874)	(3.387)	(4.137)
Black effect when						
dblack = 0, vnblack = 0	1.026	2.085	1.641	5.432*	6.096*	6.331*
	(2.457)	(2.708)	(3.120)	(3.254)	(3.161)	(3.720)
dblack = 1, vnblack = 0	-1.608	-2.097	-2.760*	-0.208	-0.276	0.469
	(1.314)	(1.390)	(1.479)	(1.490)	(1.598)	(1.749)
dblack = 0, vnblack = 1	1.909	2.453	2.993	4.537*	4.868*	4.880*
	(2.097)	(2.409)	(2.855)	(2.608)	(2.559)	(2.947)
dblack = 1, vnblack = 1	-0.725	-1.729	-1.408	-1.103	-1.503	-0.983
	(2.981)	(3.332)	(3.046)	(2.108)	(2.091)	(2.403)
N	142	132	124	133	122	115
Log-Likelihood	-212.285	-186.467	-175.082	-203.166	-178.198	-167.010
Attorney Controls		✓	✓		✓	✓
Foreperson Controls			✓			✓

Notes: Robust standard errors are reported in parentheses. Sample limited to cases with more than 5 jurors responding to school, income, age, race, gender, and religion questions and to cases in which corresponding attorneys answered the voir dire satisfaction question before learning the verdict. ∗=10%, ∗∗=5%, ∗ ∗ ∗=1%. All models also control for county and case type fixed effects, and case and voir dire characteristics.

on that basis. However, it appears that there would be little justification for attorneys to form such beliefs. When we estimate the specifications in Tables 2 and 4 excluding juror characteristics that attorneys may not be able to observe or approximate accurately, the effects of juror gender are qualitatively unaffected.[48] Thus, none of these alternative specifications provides evidence that a greater proportion of women on seated juries, when used as a proxy for other characteristics, would indicate any substantial advantage for the defense.

A more straightforward intrepretation of this divergence between our previous results and the strong preference of defense attorneys for female jurors may therefore be that attorneys lack awareness of true relations between gender and individual biases or trial outcomes. This might suggest that attorneys form preferences for the gender of seated jurors based on anecdotal notions – even if they often turn out to be incorrect – that women are markedly sympathetic to defendants or are very likely to possess other traits that are correlated with such sympathies.[49]

Attorney difficulties with observing some juror characteristics may nonetheless be affecting our results through a different channel. In contrast to our findings of a strong association between juror religiousness and pro-prosecution trial outcomes and individual juror biases, attorneys do not appear to hold strong preferences for the average religiousness of the seated jury. A possible explanation for the absence of a clear link between religiousness and attorney preferences is that the religiousness of a potential juror is less likely to be observable to the attorneys than other characteristics. Hence, while attorneys may be aware of an effect of religiousness on verdicts in general, they may have only a very limited notion either of the religiousness of the seated jurors or the expected magnitude of the effect of their religiousness in any particular case.[50]

[48]Most notably, the strong tendency for female jurors to rate the prosecutor's skill highly, from column 3 of Table 2, is unchanged in size and significance across every alternative specification. Similarly, the female effect on the rating of the defense's case becomes slightly smaller in absolute magnitude and loses significance in some alternative specifications, but is always negative.

[49]Table 5 also provides weak evidence that prosecuting attorneys prefer juries with a lower proportion of females when there is a female victim. This might suggest that attorneys also rely to some extent on anecdotal and frequently mistaken impressions that women are harsher judges of female victims.

[50]One hypothesis arising from these conjectures is that religiousness may be more strongly correlated with attorney satisfaction in cases in which attorneys were able to gather greater individual information about the potential jurors through the use of questionnaires. We test this hypothesis by interacting the average level of religiousness of the jury with a dummy indicating whether questionnaires were used in voir dire. In some specifications, we find that, when a questionnaire was used, a higher proportion of religious jurors is weakly associated with lower relative satisfaction for the defense attorney compared to the prosecution, broadly in line with our findings on individual juror biases and trial outcomes. As mentioned in Section 2, regardless of whether attorneys have access to results from supplementary questionnaires, they do know the basic occupational and biographical details that courts routinely collect from all members of the jury pool upon or prior to arrival at the courthouse. We therefore consider all of the other juror characteristics we observe

Table 5 also suggests that prosecuting attorneys have a weak preference for juries with fewer years of education on average, while defense attorneys may have a very weak preference for the opposite. This corresponds with the direction of the weak effects we found of education on trial outcomes, though not with the association we found between individual juror education and a greater tendency to interpret trial evidence in favor of the prosecution. Unfortunately, the imprecision of the estimates makes attorney preferences over juror education difficult to analyze in more depth.

Finally, our results on attorney preferences for the proportion of blacks on the jury are inconclusive and sometimes in conflict with what would be expected given our findings on the impact of racial composition on verdicts. As discussed in Section 4, individual black jurors exhibit strong pro-defense biases that are only slightly more prominent when the defendant is black; while, in cases with a black victim, a higher proportion of blacks on the jury is associated with more favorable trial outcomes for the defendant regardless of defendant race. However, results in Table 5 suggest that the prosecutor holds some preference for a greater proportion of black jurors when the defendant is not black, and also appear to suggest that the defense may hold a weak preference for a less black-dominated jury when the defendant is black, regardless of victim race. These results could perhaps be interpreted, similar to those regarding juror gender, as indicating some degree of reliance on folk wisdom amongst attorneys: prosecutors may believe that black jurors are unconditionally more likely to be sympathetic towards black defendants, but simply not pay much attention to juror race when the defendant is not black. However, inferences of this nature are difficult to support, as the estimates of these race effects are imprecise, and their magnitude and significance are sensitive to alternative specifications.

Another potential interpretation of these race results is that attorneys are expressing their satisfaction relative to some unobserved benchmark: for example, as mentioned earlier, a pool of mostly black potential jurors is likely to result in most of the seated jurors also being black regardless of what the attorneys accomplish in jury selection, in which case the prosecutor might be relieved and the defense attorney disappointed with any situation except an all-black seated jury. It is also possible that, in response to increasing scrutiny of race-based removal of potential jurors during jury selection, attorneys on both sides may have moved away from treating racial composition as a predominant goal in jury selection and instead formed stronger preferences for other juror characteristics.[51] Our findings on

to also have been at least approximately known by the attorneys from their own observation and separate information sources. Correspondingly, we find little impact of juror questionnaires on the effects of the other juror characteristics on attorney preferences in alternative specifications with extended interactions.

[51] As discussed in Section 2, striking jurors based solely on race is illegal, although there are practical difficulties with enforcing this. We control for whether any Batson objections were raised in all specifications

the strong and robust effects of the jury's average income and female representation on attorney satisfaction certainly support the hypothesis that attorneys have multi-dimensional preferences across characteristics besides race. It is thus possible that any estimated race effects on attorney satisfaction are driven by a few isolated trials in our dataset, implying that the best interpretation might be that these apparent race effects are just unreliably estimated and masking a true zero effect.

In summary, we find evidence that attorneys are aware of some relations between juror characteristics and trial outcomes. Of particular note are the strong and opposing preferences over the average income of the jury held by attorneys on both sides, which indicate that the ability to correctly anticipate the role of juror income in predicting juror biases and verdicts is widespread in the legal profession. This serves to underscore the concerns that arise from the existence of these relations in the first place, as it raises the possibility – or at least the suspicion – that attorneys could use this knowledge to manipulate trial outcomes. At the same time, our findings suggest that any attorney efforts in this regard may be hampered by an imperfect understanding of the relation between some other juror characteristics and trial outcomes and by an inability to observe some characteristics.

6 Conclusion

Our results give some justification for concerns about a lack of impartiality and trustworthiness in U.S. criminal jury trials. Defendants whose trials are heard by juries with a higher average income and fewer blacks can expect the individual jurors to interpret the case and evidence less in their favor and can expect a greater likelihood of conviction. Moreover, as attorneys appear to have some ability to correctly anticipate these relations, defendants must also worry that the prosecution will attempt to push the composition of the jury in these specific directions. These findings hardly inspire confidence in the ability of the legal system to deliver fair outcomes.

However, our findings also suggest some potential silver linings. First, our results do not point to race as the primary source of bias in criminal trials. There is a popular perception that attorneys predominantly target jurors for removal on the basis of race during jury selection, resulting in widespread discriminatory exclusion of minorities from jury service and a severe bias in verdicts against minority defendants. This perception may be primarily driven by certain high-profile cases (such as the O.J. Simpson criminal trial or the first

in Table 5, but cannot observe which side raised such objections or what, if any, remedy was provided as a result. We have attempted to test for the sensitivity of attorney preferences over race to the degree of scrutiny concerning race-based strikes by interacting with the Batson indicator, but unfortunately the small sample size prevents these effects from being estimated precisely enough to draw any conclusions.

trial of the officers who beat Rodney King), but it has also received some support from various notable studies (especially Anwar et al., 2012b, Baldus et al., 2001, and Bowers et al., 2001). Our results certainly confirm that black jurors tend to hold sentiments favorable to defendants. On the other hand, our estimated effects of juror race on verdicts are less pronounced than those on individual predispositions, and less pronounced as well than our estimated effects of certain other juror characteristics on verdicts. Furthermore, our results on attorney preferences in jury selection, while difficult to interpret, do not appear to indicate that attorneys consider juror race to be associated with a substantial advantage to either side. The difference in degree between our findings and those of other studies can likely be explained by two important factors: the high and varying racial diversity in and across the counties covered by our dataset; and the fact that we are able to jointly control for such a breadth of juror characteristics. The robust income effects we find throughout our analysis suggest that this may be an especially important omitted variable in other studies, leading to an upward bias in the absolute magnitude of their estimated race effects due to the generally greater population-wide propensity for minorities to have low incomes.

A second potential reason for optimism revealed by our results is that the substantive evidentiary and legal aspects of trials remain the predominant determinants of verdicts. The jury composition effects that we identify are small compared to the effect of evidentiary strength, and account for less of the explained variation in verdicts than do the variables related to the facts and legal doctrines central to the trial. Of course, the presence of any significant jury composition effect could be interpreted as undermining the fairness and validity of jury trials in general; and in a specific trial in which a defendant is truly innocent but nonetheless faces moderately strong circumstantial evidence, even a small effect could induce a false conviction if the prosecutor manages to alter the jury composition enough.

This question of how well attorneys can leverage jury composition effects to affect verdicts is one that our present work only partially illuminates. To be able to do so at all, attorneys must of course be aware of the jury composition effects in the first place, and this points to a third potential silver lining of our results. While attorneys on both sides seem to be well aware of the effect of juror income on trial outcomes, there are also signs that they are at least partially mistaken about the effects of other characteristics. The defense's strong preference for women on seated juries, in particular, seems misguided in comparison to the effects of juror gender on verdicts and individual biases that we estimate. And whether attorneys are aware of the strong effects of religiousness or not, it appears that they have trouble observing this trait in jurors.

But the ability of attorneys to affect trial outcomes through jury composition effects also hinges on how much they can actually manage to manipulate jury composition in voir dire,

and our present work is silent on this issue. No matter how strong some individual juror biases might be and how accurately attorneys anticipate them, the success of an attorney of a given side in stacking the jury will be tempered by the actions and strategies of the opposing counsel and by the institutional constraints imposed within the jury selection process. In other words, the question of whether attorneys are aware of jury composition effects is moot if they have no freedom to make use of such awareness in voir dire or if the actions of one attorney can be perfectly offset by the other. This might suggest that a potentially attractive approach to addressing some of the concerns arising from the effects of jury composition on trial outcomes would be to simply limit attorney freedoms in voir dire. However, in Lehmann and Smith (2012), we model attorney behavior in jury selection to evaluate these questions in detail, and our findings suggests that such a policy may have its own drawbacks. In particular, we find that giving attorneys *more* freedom in voir dire can be beneficial in some cases in terms of allowing otherwise unobservable juror biases to be revealed. Having access to this additional information can allow attorneys to break their reliance on stereotypes based on observable juror characteristics like race and income, though trial outcomes will nonetheless be affected if there are asymmetries across attorneys in how well each side can access and use this information.

While we are therefore hesitant to make definitive policy recommendations at this stage, our present findings do make an important contribution to jury and broader legal reform debates by providing the most convincing evidence to date that the sources of jury bias run deeper than race. Our results suggest that these sources are nuanced and multi-faceted, encompassing several juror characteristics and attorney beliefs about them. The issue may thus require an equally nuanced and multi-faceted policy response. Exploring how our results generalize to regions and types of cases not covered by our dataset is a crucial next step in eventually formulating such a response.

References

Abrams, David S., Marianne Bertrand, and Sendhil Mullainathan (2011) "Do Judges Vary in their Treatment of Race?" *Journal of Legal Studies, forthcoming.*

Alesina, Alberto F. and Eliana La Ferrara (2011) "A Test of Racial Bias in Capital Sentencing," *National Bureau of Economic Research Working Paper*, No. 16981.

American Civil Liberties Union of Northern California (2010) "Racial and Ethnic Disparities in Alameda County Jury Pools," *Report.*

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson (2012a) "A Fair and Impartial Jury? The Role of Age in Jury Selection and Trial Outcomes," *National Bureau of Economic Research Working Paper*, No. 17887.

——— (2012b) "The Impact of Jury Race in Criminal Trials," *Quarterly Journal of Economics*, Vol. 127, pp. 1017–1055.

Baetschmann, Gregori, Kevin E. Staub, and Rainer Winkelmann (2011) "Consistent Estimation of the Fixed Effects Ordered Logit Model," *Institute for the Study of Labor Discussion Paper*, No. 5443.

Baldus, David C., George Woodworth, David Zuckerman, Neil Alan Weiner, and Barbara Broffitt (2001) "The Use of Peremptory Challenges in Capital Murder Trials: A Legal and Empirical Analysis," *University of Pennsylvania Journal of Constitutional Law*, Vol. 3, pp. 3–170.

Bowers, William J., Benjamin D. Steiner, and Marla Sandys (2001) "Death Sentencing in Black and White: An Empirical Analysis of the Role of Jurors' Race and Jury Racial Composition," *University of Pennsylvania Journal of Constitutional Law*, Vol. 3, pp. 171–274.

Diamond, Shari S., Destiny Peery, Francis J. Dolan, and Emily Dolan (2009) "Achieving Diversity on the Jury: Jury Size and the Peremptory Challenge," *Journal of Empirical Legal Studies*, Vol. 6, pp. 425–449.

Equal Justice Initiative (2010) "Illegal Racial Discrimination in Jury Selection: A Continuing Legacy," *Report.*

Fields, Gary S. (2004) "Regression-Based Decompositions: A New Tool for Managerial Decision-Making," *Cornell University Working Paper.*

Ford, Roger Allen (2010) "Modeling the Effects of Peremptory Challenges on Jury Selection and Jury Verdicts," *George Mason Law Review*, Vol. 17, pp. 377–381.

Gobert, James J., Ellen Kreitzberg, and Charles H. Rose (2009) *Jury Selection: The Law, Art, and Science of Selecting a Jury*: West.

Greene, W. H. and D. A. Hensher (2010) *Modeling Ordered Choices: A Primer*: Cambridge University Press.

Grosso, Catherine M. and Barbara O'Brien (2012) "A Stubborn Legacy: The Overwhelming Importance of Race in Jury Selection in 173 Post-*Batson* North Carolina Capital Trials," *Iowa Law Review*, Vol. 97, pp. 1531–1559.

Hannaford-Agor, Paula L., Valerie P. Hans, Nicole L. Mott, and G. Thomas Munsterman (2002) "Are Hung Juries a Problem?" *National Center for State Courts Report*.

——— (2003) "Evaluation of Hung Juries in Bronx County, New York, Los Angeles County, California, Maricopa County, Arizona, and Washington, DC, 2000–2001," *National Center for State Courts User Guide. Williamsburg, VA: National Center for State Courts [producer]. Ann Arbor, MI: Interuniversity Consortium for Political and Social Research [distributor]*.

Hoffman, Morris B. (2006) "Unnatural Selection," *New York Times, March 6*.

Iyengar, Radha (2011) "Who's the Fairest in the Land? Analysis of Judge and Jury Death Penalty Decisions," *Journal of Law and Economics*, Vol. 54, pp. 693–722.

King, Gary (1989) "A Seemingly Unrelated Poisson Regression Model," *Sociological Methods and Research*, Vol. 17, pp. 235–255.

Kressel, N. J. and D. F. Kressel (2002) *Stack and Sway: The New Science of Jury Consulting*: Westview Press.

Lee, Jean (2010) "Do Jurors Discriminate? Evidence from State Juror Selection Procedures," *5th Annual Conference on Empirical Legal Studies Paper*.

Lehmann, Jee-Yeon K. and Jeremy Blair Smith (2012) "Attorney Empowerment in Voir Dire and the Racial Composition of Juries," *Working Paper*.

Lieberman, J. D. and B. D. Sales (2007) *Scientific Jury Selection*: American Psychological Association.

Linder, Douglas (2007) "The Trials of Los Angeles Police Officers' in Connection with the Beating of Rodney King," *Social Science Research Network Working Paper*.

Mize, Hon. Gregory E., Paula Hannaford-Agor, and Nicole L. Waters (2007) "The State-of-the-States Survey of Jury Improvement Efforts: A Compendium Report," *National Center for State Courts, Charlottesville, VA*.

Myers, David G. and G. D. Bishop (1970) "Discussion Effects on Racial Attitude," *Science*, Vol. 169, pp. 778–779.

Neilson, William S. and Harold Winter (2000) "Bias and the Economics of Jury Selection," *International Review of Law and Economics*, Vol. 20, pp. 223–250.

Norton, M. I., S. R. Sommers, and S. Brauner (2007) "Bias in Jury Selection: Justifying Prohibited Peremptory Challenges," *Journal of Behavioral Decision Making*, Vol. 20, pp. 467–479.

Pfeifer, J. E. (1990) "Reviewing the Empirical Evidence on Jury Racism: Findings of Discrimination or Discriminatory Findings?" *Nebraska Law Review*, Vol. 69, pp. 230–250.

Posey, A. J. and L. S. Wrightsman (2005) *Trial Consulting*: Oxford University Press.

Rose, M. R. (1999) "The Peremptory Challenge Accused of Race or Gender Discrimination? Some Data from One County," *Law and Human Behavior*, Vol. 23, No. 6, pp. 695–702.

Rothwax, Harold J. (1996) *Guilty: The Collapse of Criminal Justice*: Random House.

Shayo, Moses and Asaf Zussman (2011) "Judicial Ingroup Bias in the Shadow of Terrorism," *Quarterly Journal of Economics*, Vol. 126, pp. 1447–1484.

Sommers, Samuel R. (2008) "Determinants and Consequences of Jury Racial Diversity: Empirical Findings, Implications, and Directions for Future Research," *Social Issues and Policy Review*, Vol. 2, pp. 65–102.

Sommers, Samuel R. and Phoebe E. Ellsworth (2003) "How Much Do We Really Know about Race and Juries? A Review of Social Science Theory and Research," *Chicago–Kent Law Review*, Vol. 78, pp. 997–1031.

Starr, V. Hale and Mark McCormick (2001) *Jury Selection, Third Edition*: Aspen Publishers Online.

Sunstein, Cass R. (2000) "Deliberative Trouble? Why Groups Go to Extremes," *The Yale Law Journal*, Vol. 110, pp. 71–119.