Seen to Be Done: A Graphical Analysis of Peremptory Challenge

Chris Salahub

December 29, 2019

Abstract

The legal practice of peremptory challenges is described, outlining its past and present racial controversies as well as the modern defences typically provided in its favour. These arguments are analyzed statistically using novel visual tools including the mobile plot and the positional boxplot, which were developed to explore the impact of race on the exercise of peremptory challenges in several data sets (Wright et al. (2018), Grosso and O'Brien (2012), and Baldus et al. (2001)). Mulitnomial regression models motivated by these visualizations are fit and used to generate precise parameter estimates which indicate the dominance of race in peremptory challenge decisions for venire members in the Wright et al. (2018) data. Trial level summaries of the data from Wright et al. (2018) are produced and discussed in the context of the results from the venire member models.

1 Introduction

Recent racial controversy (Harris (2018); MacLean (2018)) related to the use of the peremptory challenge in *R. v. Stanley* (Quenneville and Warick (2018b)) have culminated in legislative changes to the legal practice in Canada with the proposition of Bill C-75 by the 42nd Parliament of Canada (42nd Parliament of Canada (2018a)). With Bill C-75 currently moving through the Canadian parliamentary system, having completed its second reading in June 2018 (42nd Parliament of Canada (2018b)), an evaluation of the practice of peremptory challenge is warranted. A great deal of ink has already been spilled on both sides of the debate (Hasan (2018); Zinchuk (2018); Roach (2018)), but startlingly little of this discussion has been based on any hard, quantitive evidence on the impact of peremptory challenge in jury selection. This paper aims to provide analysis and evidence to illuminate the topic further by analyzing three separate peremptory challenge data sets collected in the United States, namely the data from Wright et al. (2018), Grosso and O'Brien (2012), and Baldus et al. (2001), henceforth the "Sunshine," "Stubborn," and "Philadelphia" data sets respectively. While this data cannot reveal anything about the alleged racial motivation of peremptory challenge use in *R. v. Stanley*, a wider view of the practice is a more sober place to assess its role in modern jury trials than the dissection of a particular controversial case.

Of course, this work is not the first such investigation. Wright et al. (2018), Grosso and O'Brien (2012), and Baldus et al. (2001) have performed analysis on the factors which impact the use of peremptory challenges in their respective data sets. All of these investigations indicated that race

was an important factor in determining if a venire member was struck. Numerous others have performed unique legal, empirical, and analytical analyses of the jury selection process, including Hoffman (1997), Van Dyke (1977), Hans and Vidmar (1986), Brown et al. (1978), and Ford (2010). Most of the authors which have performed such analysis arrive at similar conclusions on the general importance of race in the exercise of peremptory challenges, and the negative impact this has on the operation and perception of justice in the legal system. Hoffman (1997) gives an exceptionally negative analysis of peremptory challenges from a legal perspective, while the game theory analysis of Ford (2010) suggests that the use of peremptory challenges may even be counter-productive.

What is, perhaps crucially, missing from this rich analysis is an effective method of communicating these results. While the tables generated to summarize the previous analyses certainly contain all the data necessary to evaluate strike patterns, they fail to be accessible to a casual reader, as they require some degree of commitment and focus to interpret and compare. Visual representations of the data which could be used for such quick comparison and interpretation would facilitate dissemination of the empirical results of these analyses to a broader audience, and would make the work of comparing and interpreting data sets far more intuitive than the current table representations. This work endeavours to provide such visual tools.

Consequently, this work proceeds in four parts. Section 2 provides the necessary legal context to understand the motivation of the previous investigations. In 2.1, the general jury selection procedure is presented before the modern controversies of this process are outlined in 2.2. Legal arguments for both the jury and the peremptory challenge are provided interspersed in this modern history in **??** and **??**. With the necessary context provided, Section 3 proceeds to discuss the three data sets obtained, explaining the sources and collection methods before detailing cleaning and preprocessing. Section 4 then provides the details and results of the analysis performed on the different data sets. It begins by performing statistical analysis of one common argument in favour of peremptory challenge in 4.1 before visualizing the Sunshine data in 4.2 and 4.3. Mobile plots are the primary tool used for this visual analysis of the data, and every visualization of the Sunshine data set is compared to analogous visualizations of the Stubborn and Philadelphia data sets. The implications of their similarities for generalization are discussed. These visual analyses are then used to motivate model selection in 4.4 in order to estimate more precisely the impact of race in the Sunshine data. These results and findings are summarized in Section 5. Recommendations based on the observations obtained are provided alongside suggestions for future work.

2 Peremptory Challenges and Previous Work

As the practice of peremptory challenges in a jury trial system is a highly specific procedure which may be unfamiliar to the reader, a brief exploration of the history, motivation, and current use of peremptory challenges is presented here. It is not exhaustive, but rather explains the terms used and the process of peremptory challenges generally. The references provided throughout are an excellent starting point for interested and motivated readers hoping to learn more.

2.1 Jury Selection Procedures

While the process of jury selection varies by jurisdiction and crime severity, the general steps of jury selection shared by the vast majority of jury trials are outlined below. More detail and a discussion

of the diversity of jury selection procedures can be found in Ford (2010), Hans and Vidmar (1986), and Van Dyke (1977). To select a jury:

1. Eligible individuals are selected at random from the population of the region surrounding the location of the crime using a list called the *jury roll*, the sampled individuals are called the *venire*

2. The venire is presented to the court, either all at once or sequentially (borrowing the names of Ford (2010): the "struck-jury" system and the "sequential-selection" system, respectively)

3. The prosecution and defence question the presented venire member(s) in a process called *voir dire*, after which there are three possible outcomes for each venire member:

   (a) The venire member is removed with cause, the cause provided by either the prosecution or defence and admitted by the judge

   (b) The venire member is removed by a *peremptory challenge* by the prosecution or defence, where no reason need be provided to the court; such privileged rejections of a venire member are limited in number for both lawyers (in Canada a maximum of 20 such challenges per side per defendant are allowed [Government of Canada (1985)])

   (c) The venire member is accepted into the jury, and so becomes a juror

4. Steps i-iii are repeated until the desired number of venire members have been accepted into the jury, typically 12.

A great deal of variation in the jury selection process is excluded from these general steps (Van Dyke (1977); Hans and Vidmar (1986)), but only some of these regional differences impact the function of peremptory challenges. While the exercise of peremptory challenges occurs both under the struck-jury and sequential-selection systems of voir dire, Ford (2010) and Van Dyke (1977) note that the predominant method in the United States and Canada is the sequential-selection system. For a more detailed discussion of the history and philosophy of this system, see von Moschzisker (1921), Hoffman (1997), Woolley (2018), Supreme Court of Canada (1991), Hans and Vidmar (1986), and Van Dyke (1977)

2.2 Modern Peremptory Challenge Controversy

The use of peremptory challenges in modern jury systems has, perhaps, generated more controversy than any other aspect of jury trials. The privileged removal of a venire member without any justification has seen persistent allegations of abuse, often around the use of these challenges by state prosecutors.

In the United States, the criticism has focused on racial discrimination and has led to significant changes in their allowed use through cases such as *Swain v. Alabama* (Supreme Court of the United States (1965)) and *Batson v. Kentucky* (Supreme Court of the United States (1986)). The first of these cases, *Swain v. Alabama*, established in 1965 that the systematic exclusion of venire members of a particular race would be unconstitutional discrimination under the Fourteenth Amendment to the United States Constitution, but argued that a *prima facie* argument of discrimination was

not adequate to prove this[1]. This placed a significant burden on the party opposing a particular peremptory challenge to demonstrate that the specific challenge had been discriminatory.

However, this ruling was overturned only 21 years later in the 1986 case *Batson v. Kentucky*, resulting in the creation of a new challenge which could be used to nullify peremptory challenges: the so-called "Batson Challenge". This new system allowed the party objecting to a challenge to use a *prima facie* argument which must be countered by a race-neutral reason that satisfies the judge. If no such reason could be supplied, the peremptory challenge would not be allowed. While the effectiveness of this system of additional challenges is questionable both practically and in abstract (Page (2005); Morehead (1994); Hoffman (1997)), it has only been extended to allow Batson Challenges for both the sex and race of venire members [Supreme Court of the United States (1993)].

Racial controversies have also been present in Canada before *R. v. Stanley*. Racial bias against First Nations venire members in Manitoba was alleged in 1991 in a report produced by a provincial inquiry into peremptory challenge use (Roach (2018)). More damning was the Iacobucci Report on First Nations representation in juries. This report proposed an explicit restriction to the practice when it recommended " an amendment to the Criminal Code that would prevent the use of peremptory challenges to discriminate against First Nations people serving on juries."

Despite the legal changes and recommendations, peremptory challenges are defended as a key component of the jury selection process by some. The modern defence is perhaps best described by Justice Byron R. White in Supreme Court of the United States (1965):

> The function of the challenge is not only to eliminate extremes of partiality on both sides, but to assure the parties that the jurors before whom they try the case will decide on the basis of the evidence placed before them, and not otherwise. In this way, the peremptory satisfies the rule that, "to perform its high function in the best way, justice must satisfy the appearance of justice."

Such a justification is reminiscent of the now famous words of Lord Chief Justice Hewart in *R. v. Sussex Justices* in 1924: "Justice should not only be done, but should manifestly and undoubtedly be seen to be done". While these words originally only referred to the pecuniary interest of court staff involved in the case, they have since come to express the idealized expectation that both the defence and prosecution find the judge and jury acceptable [Richardson Oakes and Davies (2016)]

This defence suggests two modern justifications for the peremptory challenge. The first is that it is necessary to remove venire members with "extreme" bias, and the second is that it creates a jury which is composed of jurors mutually acceptable to both the defence and the prosecution. Those who defended the peremptory challenge in Canada after *R. v. Stanley*, including Hasan (2018) and Macnab (2018), seem to use this defence or some variant of it to argue in favour of keeping the practice.

[1] In the actual case, not a single black juror had sat on a jury in Alabama in the previous 15 years, despite composing 26% of the jury-eligible population. In Swain's trial, six of the eight black venire members were rejected by state prosecutor peremptory challenges, and the other two removed for cause, leaving not a single black juror to judge Swain, a black man. This was the *prima facie* argument presented by Swain's defence team against the state prosecutors of Alabama.

2.3 Previous Invesigations

These arguments have already stoked academic investigation of the practice of peremptories. Legal analyses have been presented by many, such as Hoffman (1997), Broderick (1992), and Nunn (1993), and the large majority of these analyses take a negative view of the peremptory challenge as it currently stands. They typically either recommend large modifications to the practice beyond the Batson Challenge or the abolition of the practice altogether.

These legal analyses have been complemented by theoretical explorations by Ford (2010) and Flanagan (2015) using game theory. Both investigations indicate that the current system of peremptory challenges may produce juries which are, counterintuitively, biased towards conviction or acquittal with a high proportion of extremely biased members of the population. This implies that peremptory challenges are more useful for the purpose of "stacking" a jury in favour of one side[2].

Hoffman (1997) provides a rigorous counter-argument to the assertion that peremptory challenges ensure a jury both the prosecution and defence accept. The reasoning supporting this argument fails to account for the impact of removing an unbiased juror to both the perception of justice and the composition of the final jury. Rather, it focuses singularly on the inclusion of a biased juror as the cause of a contentious jury. Such a narrow view cannot realistically be held in light of the decisions of *Batson v. Kentucky* and *J.E.B. v. Alabama*, which implicitly acknowledge the corrosive nature of unjustified strikes to the core principles of an unbiased jury of peers.

Even more pertinent to this work are the empirical analyses performed in Baldus et al. (2001), Wright et al. (2018), Grosso and O'Brien (2012), Baldus et al. (2012), and others. These have universally found illicit factors such as race to be statistically significant in the exercise of peremptory challenges. This is both in aggregate and when possible confounding factors are controlled using, for example, logistic regression. It is possible this observed discrimination is a manifestation of an inability of lawyers to articulate the specific biases they detect, though this is a weak argument given that argumentation is the speciality of the legal profession.

Lacking in all of these investigations are visualizations which can communicate the results to broader audiences than the field of legal academics. The typical presentation of empirical results utilises extensive and complex tables which require considerable mental effort to fully interpret for even experienced analysts. This stymies the communication of results and comparison of patterns between different analyses, limiting the ability of all to construct a cohesive picture of the illicit patterns present in peremptory challenge use.

3 Data

Without data, performing an analysis that incorporated more than the legal argumentation presented in Section 2 is impossible. While the motivation of this work was a Canadian case, no comprehensive data sets examining jury selection in Canada could be found. The prominence of the jury selection process in the United States garnered a more fruitful search. The author is heavily indebted to Wright et al.; Grosso and O'Brien; and Baldus et al.. These authors shared their data freely with the author, providing him with a wealth of data to analyse.

[2] In Chapter 6 of Hans and Vidmar (1986), the "science" of using peremptory challenges to construct a biased jury is described in great detail for the case of *M.C.I. Communications v. American Telephone and Telegraph*.

3.1 Jury Sunshine Project

The Jury Sunshine Project data (Wright et al. (2018)), generously provided by Dr. Ronald Wright, is the most extensive data set of the three. It contains jury data for almost all felony trial cases in North Carolina in the year 2011, providing simple demographic characteristics and trial information for 29,624 individuals summoned for jury duty in 1,306 trials. The relevant case data recorded was the presiding judge, prosecutor, defence lawyer, defendant, venire members, charges, verdict, and sentence. For venire members, the collected data included the "disposition" – i.e. removed with cause, removed by peremptory challenge, or retained on the jury – of the venire member and the party which challenged in the peremptory case. Using public voter databases, bar admission records, and judge appointment records, the data also included race, gender, and political affiliation data for the venire members, lawyers, defendants, and judges.

3.2 Stubborn Legacy Data

Grosso and O'Brien (2012) also provided data to the author, albeit a more limited set. This study, also based in North Carolina, focused on the trials of inmates on death row as of July 1, 2010, yielding a total of 173 cases. As in the Jury Sunshine study, case files and juror questionnaires were used to collect information about the court proceedings such as the peremptory challenges, venire members, presiding judge, prosecutor, and defence lawyer. Unlike the Jury Sunshine study, detailed verdict and charge information was not collected, as the pre-selection criteria of death row inmates made the verdict clear, and the death penalty can only be applied for a limited number of serious crimes.

Another critical difference between the Stubborn data and the Jury Sunshine data is the exclusion of venire members removed with cause in the Stubborn data. This was motivated by the analysis of the study, which focused on the differences between defence and prosecution peremptory challenges. Less critically, the Stubborn data set lacks data on political affiliation, which serves as a barrier to the comparison of this data to the Sunshine data on an identical basis. However, the racial data for the two is recorded in a very similar way, so this variable can, at least, be compared.

3.3 Philadelphia Data

Baldus et al. (2001) presents a similar data set to Grosso and O'Brien (2012) collected using similar means. Court files such as the juror questionnaire, voter registration, and census data were all used to complete juror demographic information for 317 venires consisting of 14,532 venire members in Philadelphia capital murder cases between 1981 and 1997. It should be noted that this data included only those jurors kept or peremptorily struck, venire members struck for cause were not included. The procedure used to determine race using the census and voter registration polls was quite complicated, but was rigorously performed using accepted census methods to a standard of 98% reliability. This data had a number of departures from the Sunshine and Stubborn data. It lacked racial information as detailed as either and recorded no information about political affiliation, futher limiting the potential for direct comparison between all three data sets.

4 Analysis

All code can be found at the author's GitHub: github.com/Salahub/peremptory_challenges.

As noted in 2.2, extensive empirical analysis of this subject and theoretical explorations have answered a great deal of questions already. It seems clear that race is an important factor in the exercise of peremptory challenges to the detriment of jury composition. Theoretical investigations have suggested that the exercise of peremptories may increase the proportion of extreme jurors rather than reduce them.

In light of this, it is natural to wonder whether there is any evidence that the most common arguments posed in favour of peremptory challenge are satisfied in this data. As discussed in **??**, there are two primary arguments. The first is the argument that the peremptory challenge is necessary to remove the "extremes of partiality" present in the venire for both sides; that is to remove the most extremely biased jurors. This goal is already partially accomplished by challenges with cause, which are designed to remove those jurors with extreme bias. The second argument is the creation of a jury which is mutually acceptable to both parties in the trial.

4.1 Extremes of Partiality

Unfortunately, not much can be said about the extremity bias, as it depends on population characteristics. Suppose, as in Ford (2010) the bias of the population for a particular trial is modelled using a beta distribution, where each individual has some bias between 0 and 1, representing that individual's subjective probability that the accused is guilty before seeing any evidence. A perfectly unbiased individual would have a bias of 0.5, and so no prior assumption of innocence or guilt. Such an individual would be equally receptive to the arguments of the prosecution and defence, and so would be an ideal juror.

Under the simplistic beta model, the diversity of beta distributions make statements about the proportion of points around 0.5 impossible without additionally assuming some parameter values. Consider parametrizing a beta distribution with α and β to give the probability density function $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\ x \in [0,1]$. If $\alpha \to 0$ and $\beta \to 0$ this distribution becomes polarized, with all of its probability mass shifted to its boundaries at 0 and 1. Conversely, as $\alpha \to \infty$ and $\beta \to \infty$ the density becomes degenerate at 0.5.

Flanagan (2015), attempts to avoid the issue of parameterization by instead simply modelling the votes of the venire members conditional on all of the features and evidence of a trial. This assumption does not allow much more to be determined, however. Using this conceptualization, the best characterization of an individual with extreme bias is an individual that would vote to convict or not convict regardless of the evidence presented. Once again, knowing the proportion of such individuals present in the broader population is necessary to make statements of the validity of a certain proportion of the venire being struck. Here the relationship is more direct, as the appropriate proportion of struck venire members would simply be the proportion of such certain convictions or acquittals in the venire.

4.2 The Impact of Race

The racial controversy surrounding peremptory challenges provides one aspect of venire member rejection which may warrant further investigation. To begin, it is informative to use aggregate

marginal distributions to explore the impact of venire member race on the peremptory strike probability, as displayed in Table 1. Of particular interest is whether any race is far more likely to be struck by peremptory challenge than the others, as this would suggest that race is the target of an undue rate of strikes.

Table 1: The conditional probability of a venire member being struck peremptorily by the simplified venire member race across data sets. These values are smaller than the values presented in the extremity analysis as only the individuals which were identifiably removed by peremptory challenge are counted in this table. Regardless, the comparisons remain similar even if the unattributed removals are included. Note that the Philadelphia trial data only indicated black and non-black venire members and so only two numbers can be reported.

Data	Black	Other	White
Sunshine	0.23	0.24	0.25
Sunshine Capital	0.22	0.27	0.27
Stubborn	0.65	0.36	0.66
Philadelphia	0.67	0.68	

These probabilities are different, but not greatly so. Indeed, the trend of higher probabilities for the removal of white jurors across all data sets is perhaps counter-intuitive given the history of controversy in the United States. In any case, the small magnitude of these differences suggests at most a weak racial bias at the aggregate level, whether or not the results are statistically significant[3].

This table also demonstrates some of the drawbacks of tables, the dominant method used to display the data throughout Wright et al. (2018), Grosso and O'Brien (2012), and Baldus et al. (2001). The table, while excellent at communicating specific values, does not provide a sense of trends or patterns without careful engagement by the reader. A critical component of the communication of any analysis and its comparison to others is the ability to quickly and effectively discern patterns in the data. Consequently, the "mobile plot" for visualizating the three way relationships of categorical variables was developed, with motivation from the hierarchy of visual perception of Cleveland and McGill (1987) and inspiration from Tufte (2001). It is used in 1 to compare dispositions of venire members by their race and the defendant race.

First, a small explanation of this mobile plot. This mobile plot displays the relationship between three categorical variables: venire member race, defendant race, and disposition (whether a venire member is struck and by whom). The vertical axis corresponds to the conditional probability of a particular disposition given a race and defendant race combination[4]. Racial combinations are placed along the horizontal axis, and each combination corresponds to one horizontal black line in the plotting area. The length of these lines is proportional to the number of venire members in the data with the corresponding racial combination, and their vertical positions are the mean conditional probability of a venire member being struck for that particular combination. The dashed vertical lines, coloured by disposition, start at this mean line and extend to the observed conditional probability of the corresponding disposition for the relevant racial combination. As a consequence, this plot can be viewed as a visualization of the test of a specific hypothesis:

[3]Consider the numerical impact of a statistically significant difference of a few percent when the jury size is 12 for each trial.

[4]Generally, any three categorical variables can be displayed using a mobile plot, as the mobile plot is used to display the distribution of a categorical variable given the combinations of two others. Mathematically it displays $X_3|X_1, X_2$.

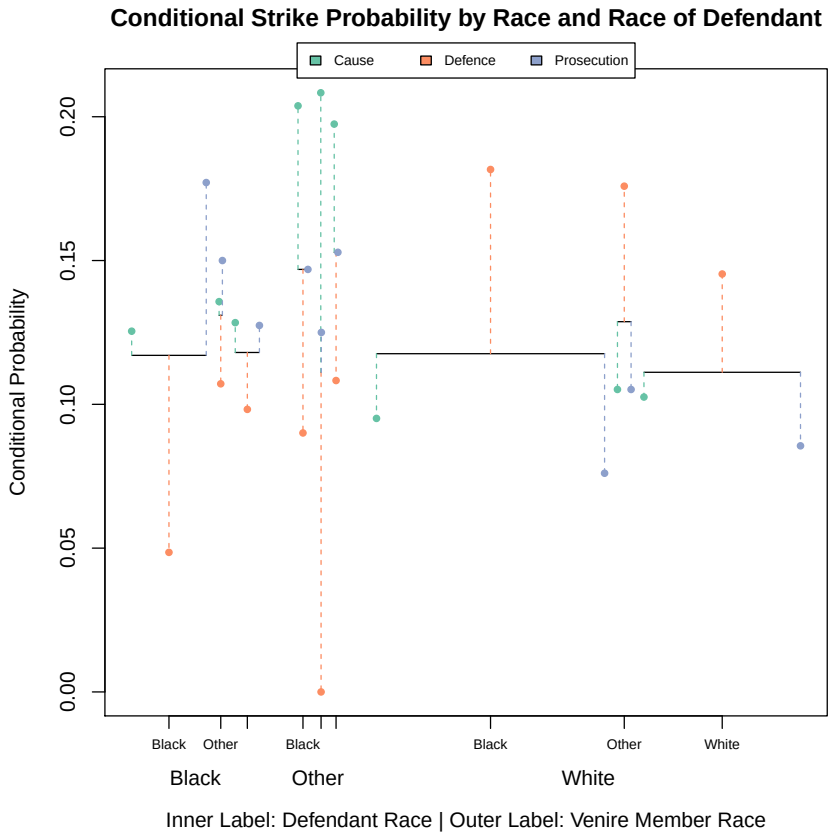


Figure 1: The conditional probability of the strike dispositions given the venire member and defendant race, with the expected value represented by the horizontal black lines, and the observed values represented by the points at the end of the dotted lines. Each horizontal black line corresponds to a particular venire member and defendant race combination, with a length proportional to the number of venire members with that combination. The dashed vertical lines, coloured by challenge source, start at these horizontal lines and end at points which show the observed probability of a venire member being struck by the source for the given racial combination.

$$D|D \in \{2,3,4\}, R, E \sim Unif(\{2,3,4\}) \tag{1}$$

Where D, R, E are random variables representing the disposition, venire member race, and defendant race, respectively. In words: the conditional distribution of the disposition given the racial combination of a struck venire member is uniform. This implies that causal challenges, defence strikes, and prosecution strikes occur with the same probability for each racial combination, though the rate may differ between racial combinations. Such a hypothesis allows for certain racial combinations to experience a higher strike rate generally, but constrains the strike rate to be the same for all parties, which would imply that all parties in the court pursue an identical strike strategy across all venire member and defendant race combinations.

Clearly, Figure 1 casts some doubt on this hypothesis. While the horizontal black lines tell a very similar story to Table 1, with little variation between them except for in small population subsets, a number of other striking patterns are visible. The first, and most obvious of these, is the main effect of venire member race. While the aggregate removal rates do not seem to depend on the race of the venire member, it is clear that the defence and prosecution pursue radically different strate-

gies. The defence seems biased towards a jury with more venire members from racial minorities. All orange points are below the horizontal lines for the black and other venire members, indicating these groups are less likely to be struck by the defence than expected, while the points are above the lines for the white venire members, indicating a higher than expected probability of defence strike for white venire members. The prosecution seems to mirror this tendency, striking the white venire members at a lower rate than expected and the black venire members more often than expected. Strikes with cause seem to show less deviation from expectation for the black and white venire members, and always deviate from expectation in the same direction as the prosecution.

The addition of defendant races shows another interesting trend. It would seem that the aforementioned tendencies of the prosecution and defence are strongest for black defendants, which have the greatest departures of the conditional probabilities from the expectation. The defense and prosecution seem to have slightly more similar habits when the defendant is white, despite their opposite tendencies in all cases. Finally, it would seem that the removals with cause have tendencies similar to the prosecution, as the points representing the conditional probability of a venire member being removed with cause are always on the same side of the expected line, an event which would occur with probability $2^{-9} \approx 0.002$ under the hypothesis of independent uniform strike rates. Further discussion of the agreement of these two strike tendencies can be found in 4.4.

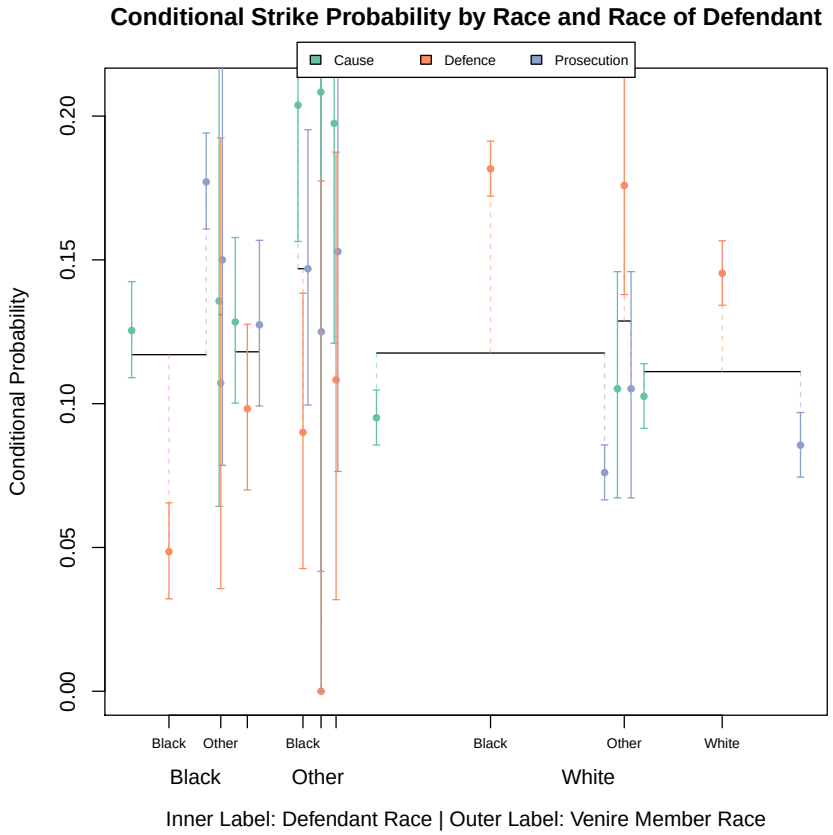


Figure 2: The plot of conditional strike probability by racial combination from Figure 1 with confidence intervals added. Note that many of the seemingly striking departures seen are insignificant when these confidence intervals are applied, especially for races other than black and white.

While Figure 1 is quite suggestive, the widths of certain horizontal black lines, in particular those for venire members with a race other than white or black, suggest that perhaps some of the more extreme tendencies are simply a result of the well-known higher variation of samples with small sizes. In order to see the true nature of the noted departures some incorporation of the variation one expects from each observed value is required. This is accomplished by the addition of approximate 95% simultaneous multinomial confidence intervals using the `MultinomialCI` package in R, which implements simultaneous confidence intervals for multinomial proportions following the method presented in Sison and Glaz (1995). These confidence intervals can be seen in Figure 2.

As suspected, some of the results for the smaller sample sizes do not seem to be significant. The results for the larger groups, in particular for white venire members or black defendants, are significant, however. It should be noted that these simultaneous confidence intervals do not constitute a proper statistical test of the impact of race, they are rather a way of visually providing a viewer some sense of the expected variability in the data over repeated sampling. More rigorous testing requires controlling for the impact of confounding factors, as done by the modelling in 4.4.

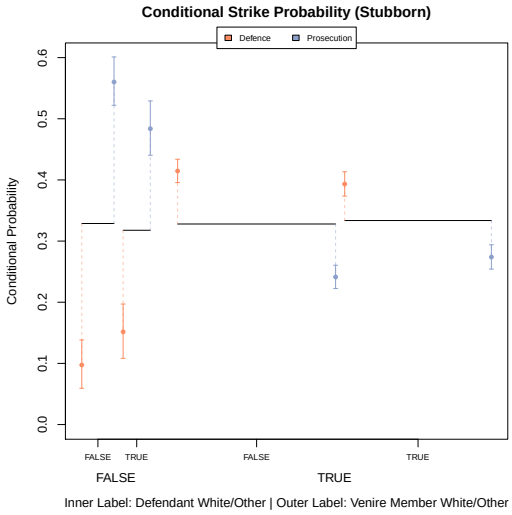
4.2.1 In the Stubborn and Philadelphia Data

Already the utility of the mobile plot, and visualizations of the data in general, should be clear. A wealth of information is displayed very simply in Figure 2. However, the real power of this plot comes with the ability to quickly compare different data sets. Consider, for example, comparing the results of Baldus et al. (2001) and Grosso and O'Brien (2012) using Table 1. In order to compare, the rows for each data set must be viewed and the numbers committed to memory before the reader moves to the appropriate row to compare values. While the simple four row and three column structure of this particular table make this rather straightforward, it becomes more difficult and less clear as these tables grow in complexity. Compare this with a cursory glance at Figure 3, which makes the similarities of the data sets immediately clear[5].

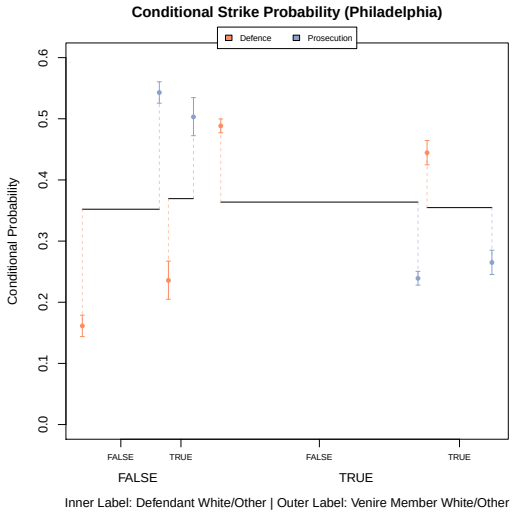
Despite the very different study populations of these three data sets, all display identical patterns, with only the magnitudes of the strike rates differing. The mobile plot format makes several interesting aspects immediately clear. The similar level of all black lines within each plot shows that in each data set, the aggregate probability of removal is similar across racial combinations, as was implied by Table 1. However, these aggregate similarities hide the vastly different strategies of the defence and prosecution which are consistent across all data sets. The defence has a tendency to retain black venire members, striking them at a lower rate than the other venire members, while the prosecution shows a pattern which mirrors that of the defence, removing more black venire members and fewer other venire members. In all data sets the gap between these probabilities and the expected strike rates is greatest for the black venire members in cases with black defendants.

It should be noted that the Sunshine data set looks most unique of the three, and this may be a result of the sample size. While the Philadelphia data and Stubborn data both collected data which

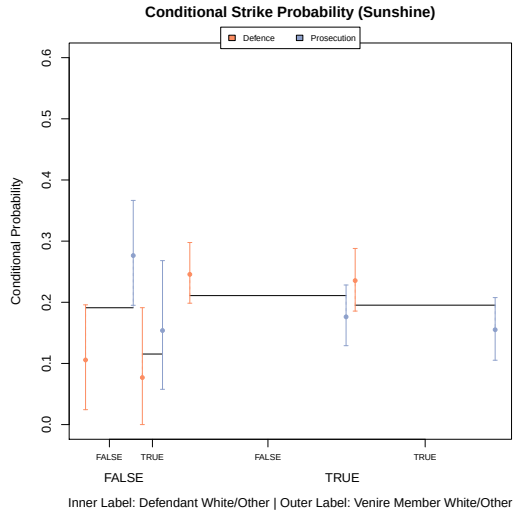
[5]Three important differences between Figure 3 and Figure 2 must be noted when interpreting these visualizations. The first of these is that data differences have limited the scope of the comparison to simply the prosecution and defence strike rates, as the Stubborn data does not include any information on strikes with cause. Second, the Sunshine data used to generate the mobile plot in Figure 3 is filtered to only the first degree murder trials, as the other two data sets only addressed capital trials, while the Sunshine data had a broader scope. Finally, the race and defendant race were further simplified to logical indicators of whether the race variable was black or not.



(a) Stubborn strike pattern



(b) Philadelphia strike pattern



(c) Sunshine strike patternx

Figure 3: The conditional probability of defence and prosecution peremptory challenge by race across all capital trials in all data sets. The pattern, though sometimes different in magnitude, is quite consistent across the three, despite the significant differences in the respective study sample universes.

included multiple years, the Sunshine data was restricted to trials which occurred in 2011. This small sample is the reason for the large confidence intervals present in Figure 3c. This does not explain the overall lower strike rate observed in the capital trials in this data, which is also visible in Table 1. This departure may be of interest for further investigation.

4.3 Other Factors

Of course, it would be incorrect to conclude immediately that the cause of the racial patterns observed across these data sets is race itself. There may be a plethora of attitudes associated with

race that could serve as legitimate cause for a peremptory challenge. As noted by Justice Byron R. White in the majority opinion in Supreme Court of the United States (1965)

> [The peremptory challenge] is no less frequently exercised on grounds normally thought irrelevant to legal proceedings or official action, namely, the race, religion, nationality, occupation or affiliations of people summoned for jury duty. For the question a prosecutor or defense counsel must decide is not whether a juror of a particular race or nationality is in fact partial, but whether one from a different group is less likely to be.

This quote leads directly to the heart of the problem. Without detailed transcripts indicating how the venire members were questioned, it cannot be known if the aggregate pattern of removal is the result of racially based strikes, or whether the lawyers determined valid reasons for a peremptory challenge during the voir dire process which are simply related to race. For example, if defence attorneys reasonably assumed that deference to authority would make a venire member insurmountably predisposed to reject their arguments about mishandling of evidence by police, this would be reasonable grounds for peremptory challenge. If such opinions were distributed heterogeneously by race, the aggregate pattern may appear to reflect racially-based decision making by the defence attorneys despite the true explanation being legitimate and non-racial.

4.3.1 Political Affiliation in the Sunshine Data

Revesz (2016) provides, inadvertently, data which might support the above defence of peremptory challenges on the basis of confounding variables. He notes that the distribution of political affiliation in the United States is not consistent across races, with black voters far more likely to vote for the Democratic Party and far less likely to vote for the Republican Party, while white voters are more uniformly distributed and likely to vote independently. If political affiliation is used as a surrogate for perspective, this suggests that the observed pattern in the Sunshine data could be the result of the defence removing conservative venire members and the prosecution removing liberal ones. As the Sunshine data has political affiliation, the results of Revesz can be compared to this data. The resulting figure, Figure 4, displays the conditional probability of political affiliation across races and genders.

What is immediately apparent viewing this plot and the data in Revesz (2016) is how closely the two data sets agree. Black voters vote Democratic a vast majority of the time, while the political leanings of white voters are less cohesive. Note additionally that the horizontal lines in this plot are all very close to 0.33, as almost all voters are Democrat, Republican, or independent. The "ideological imbalance" of racial voting tendencies, as Revesz aptly calls it, is a clear confounding factor and a possible source of a legitimate cause for an initially suspect overall trend. As such, the mobile plot was used to investigate the relationship between race, political affiliation, and disposition.

To control for the defendant race as well, which appears important in Figure 1, the venire members were split into the racial groups black, white, and other. Then mobile plots of the conditional strike probabilities for the different venire member races given the defendant race and political affiliation were generated. Figures 5a, 5c, and 5b display these mobile plots.

This sequence of three plots immediately suggests that a political argument is insufficient for this data. For each venire member race and defendant race, the political affiliation of the venire member does not radically change the pattern of strikes for any party in the court. Rather, the court tendencies for each political affiliation, venire member race, and defendant race seem to follow the

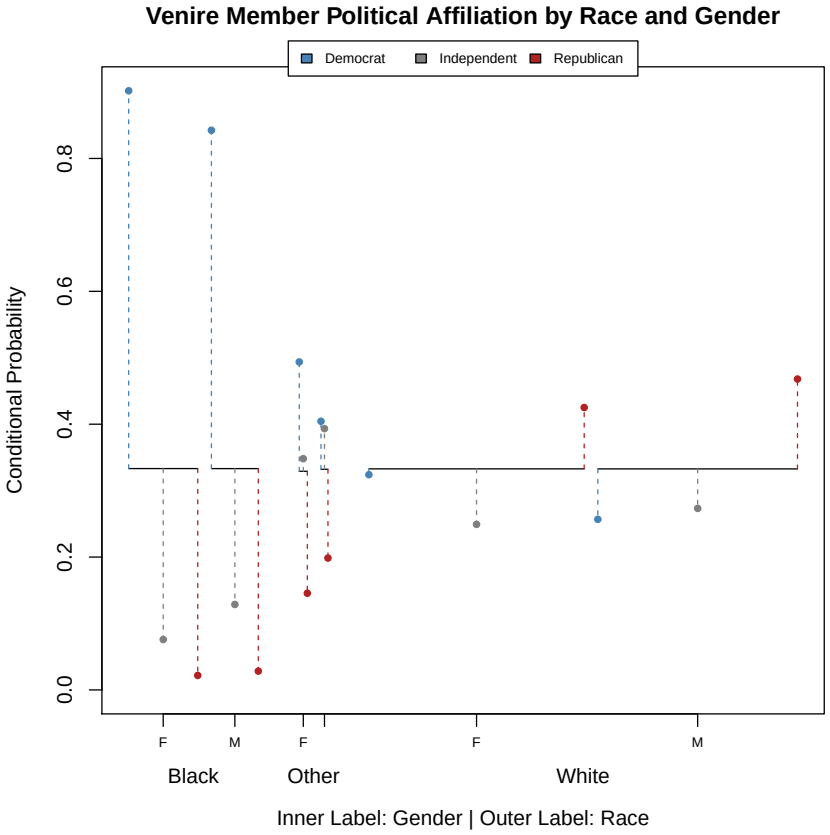


Figure 4: Conditional probabilities of political affiliation by race and gender. In this plot, it is clear the black venire members are far more homogeneous than the white venire members for both genders, with the vast majority voting Democrat.

pattern seen in 1 for all political affiliations with the exception of some very small subgroups of non-white venire members[6]. While the small sizes of many of these subgroups make this plot by no means conclusive, they suggest that a political defence of the use of peremptory challenges is inadequate to explain the differences in prosecution and defence strategy in the Sunshine data set.

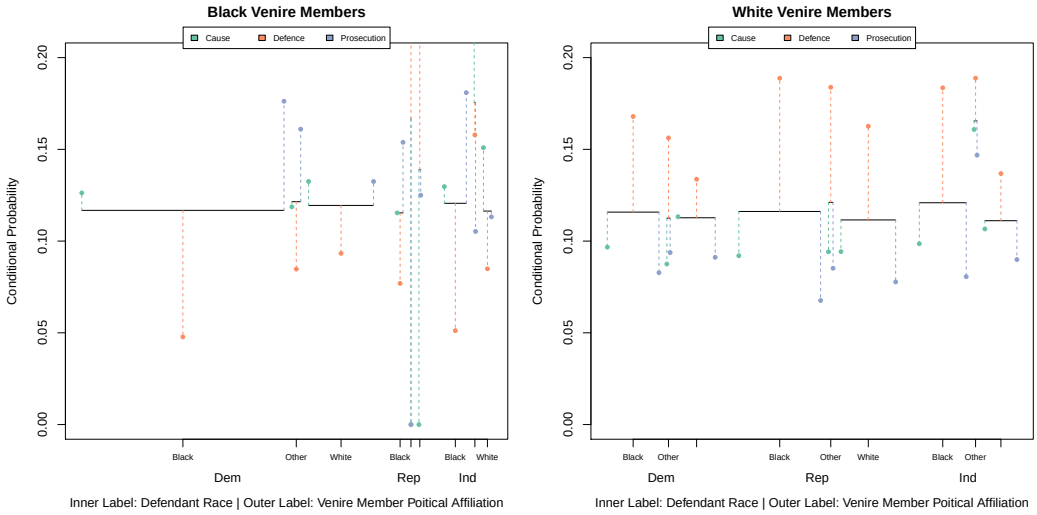
4.3.2 Gender in the Sunshine Data

Another factor which may have an impact is that of gender. While *J.E.B. v. Alabama* [Supreme Court of the United States (1993)] has also ruled peremptory challenges for this reason alone unconstitutional in the United States, it is noted in Van Dyke (1977) on page 152-153 that prosecutor's guidelines have, in the past, recommended using peremptory challenges to remove female venire members[7], and so perhaps it is a relevant factor. Additionally, there is a relationship between gender

[6]The subgroups, black republican venire members for white or other race defendants, have sizes 1 and 10, respectively. The study examines a total of 29634 venire members.

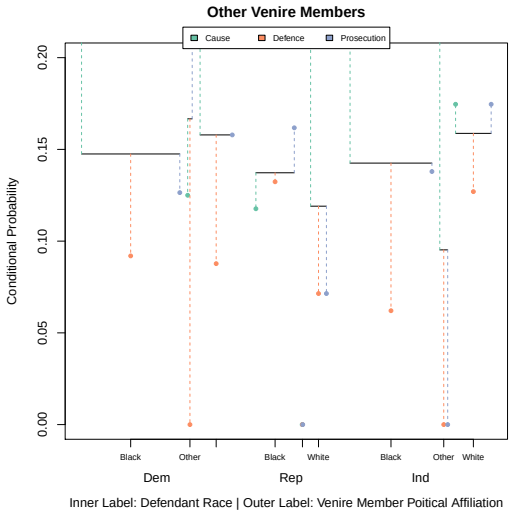
[7]Van Dyke provides specific guidelines for sex written by Jon Sparling, an assistant district attorney in Dallas, which read:

1. I don't like women jurors because I don't trust them.



(a) Black Venire Members

(b) White Venire Members



(c) Other Venire Members

Figure 5: Conditional probability of venire member strike by defendant race and political affiliation, split by race. Note how the pattern of conditional probabilities is the same across political affiliations for the same defendant race within a particular venire member race, and this pattern is the aggregate pattern seen in Figure 1

and race in the data, as noted in Wright et al. (2018): black males are highly under-represented relative to black females. Luckily, the relationship between race, gender, and peremptory challenges can be visualized quite neatly without the need to split the plots as was done in Figure 5. Figure 6

2. They do, however, make the best jurors in cases involving crimes against children.

3. It is possible that their "women's intuition"[sic] can help you if you can't win your case with the facts.

4. Young women too often sympathize with the Defendant; old women wearing too much make-up are usually unstable, and therefore are bad State's jurors.

If data on make-up use and jury outcome is ever collected perhaps Mr. Sparling's bold claim can be tested, but in light of his first point it is better treated as prejudice.

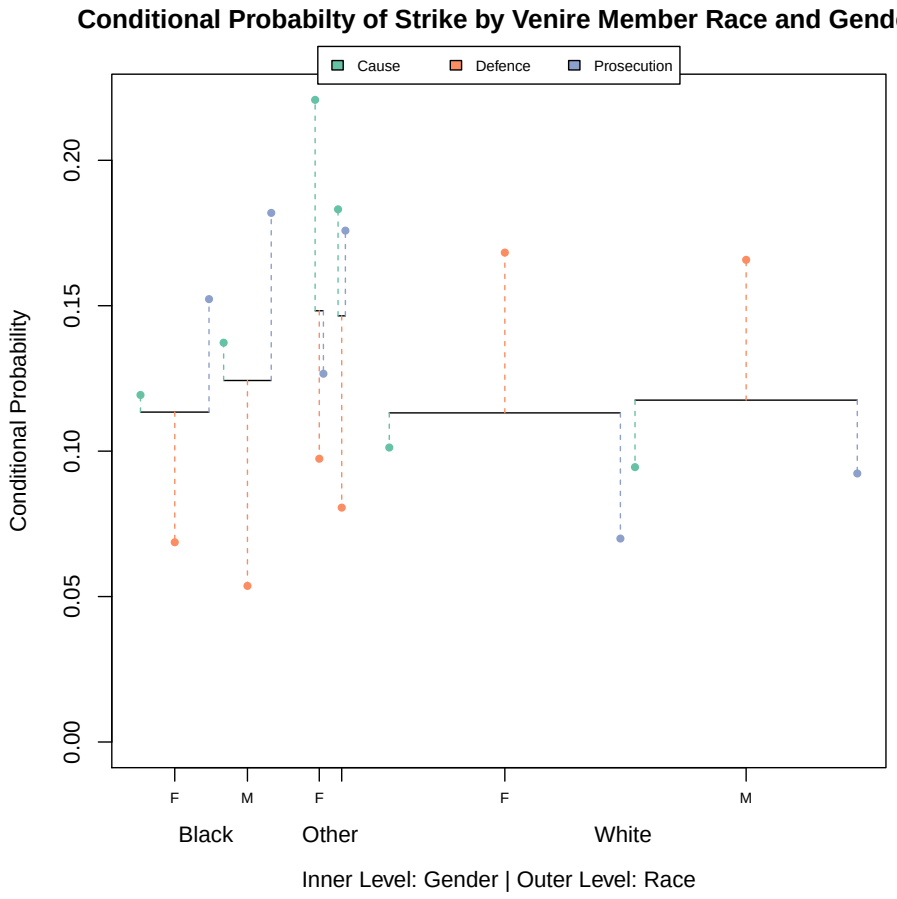


Figure 6: Conditional probabilities of strike source by race and gender. Note that the pattern is nearly identical for the genders within each racial group.

displays the mobile plot for race, gender, and striking party.

The same pattern is seen here for the most part, that of a dearth of defence strikes against visible minorities and a surplus against white venire members, with a mirrored tendency for both the prosecution and challenges with cause. One may be motivated next to ask whether such oppositional patterns are just a quirk of the data or mobile plots. Perhaps conditioning on gender alone, for example, would create a similar pattern of tendencies of the court. Figure 7 shows a plot of the court's strike tendencies given the gender of venire member and defendant gender to investigate this possibility.

The characteristic pattern seen throughout the other plots is not present. At this point the message from these additional mobile plots should be clear. That is, the dominant determinant of the strike probability for a venire member is likely their race. The race of the defendant does impact this somewhat, but plots across gender and political affiliation for both venire members and defendants in numerous combinations suggest that race is the dominant factor in determining the probability a venire member will be struck by the prosecution, defence, or removed by cause. This hypothesis, motivated by these plots, is examined more rigorously through the construction and testing of specific models in 4.4.

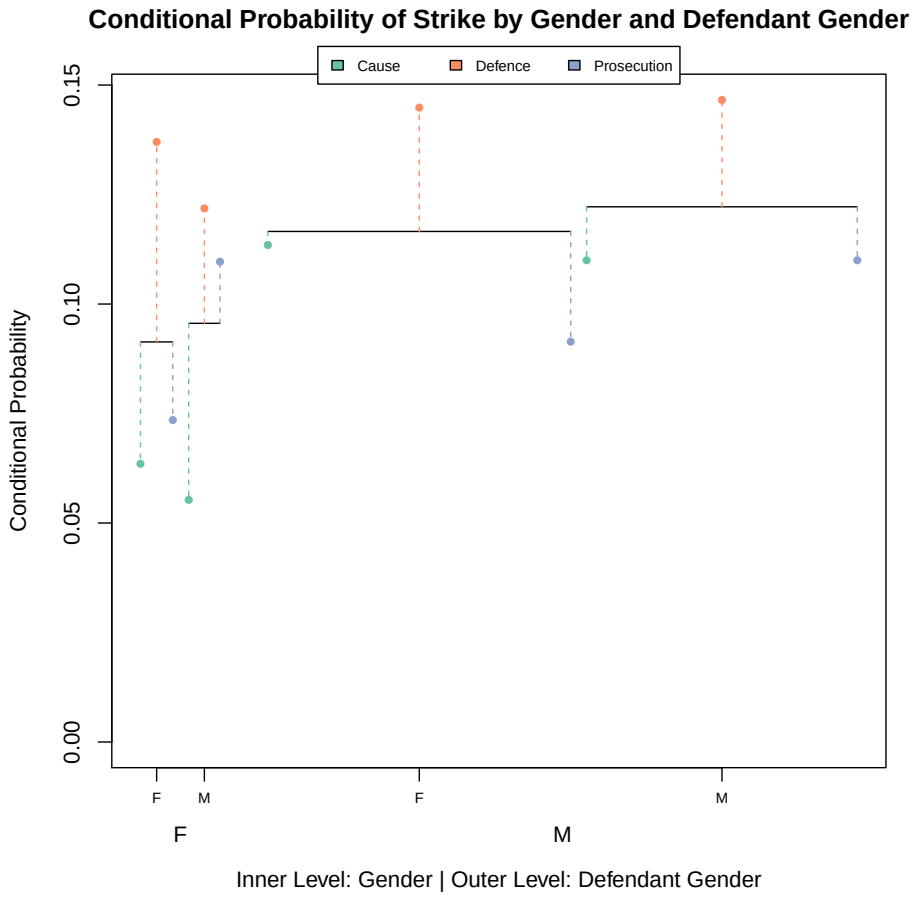


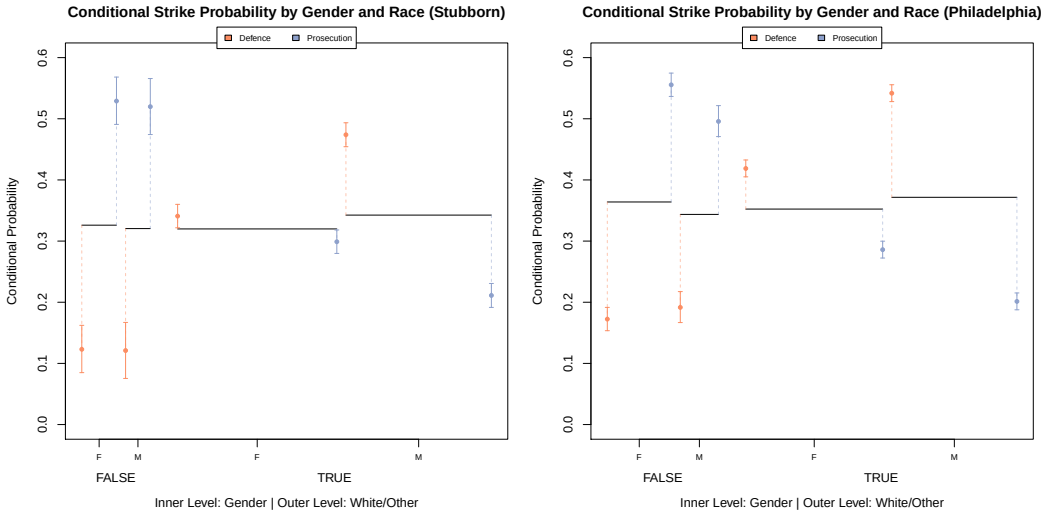
Figure 7: Conditional probabilities of striking party by venire member gender and defendant gender. Rather than the changing pattern seen with race, where the defence and prosecution show a preference for one race over others, the pattern here is rather constant, with the defence striking more than expected and the prosecution and strikes with cause generally occurring less.

4.3.3 In the Stubborn and Philadelphia Data

Grosso and O'Brien (2012) and Baldus et al. (2001) did not collect data on the political affiliation of the venire members[8], and Grosso and O'Brien (2012) did not record the gender of the defendant. Consequently, of the above plots, only Figure 6 can be compared to these two data sets. Figure 8 displays the plots. Note that as before, the data has been restricted to ensure analogous study universes across the data sets.

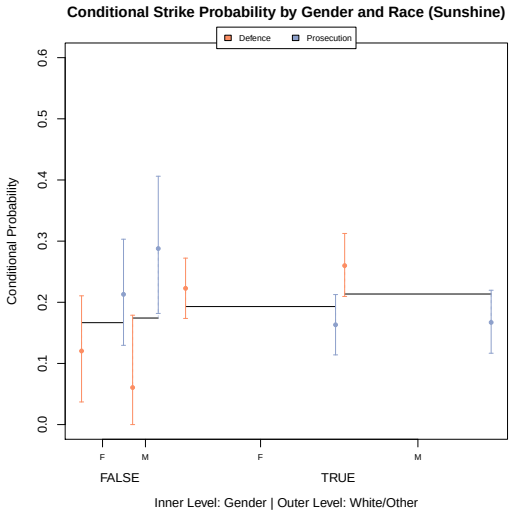
This, again, shows a pattern very similar to the overall Sunshine data, and the differences in pattern between these data sets are minor. The race of the venire member seems to be analogously important across all data sets as it was in Figure 3. Such similarity is somewhat surprising, as

[8]While it is perhaps possible that the different combinations of the attitudinal data collected by these groups could be used with data on the voter profiles most typically associated with the political affiliations of different groups in the United States, such an investigation is beyond the scope of this work. Lacking such a predictor, it is also unclear which attitudinal variables would be most comparable to the political affiliation, and so attempting to compare this pattern would be futile.



(a) Stubborn strike pattern

(b) Philadelphia strike pattern



(c) Sunshine strike pattern

Figure 8: The conditional probability of defence and prosecution peremptory challenge by venire member race and gender for all capital trials in all data sets. The pattern, though sometimes different in magnitude, is quite consistent across the three examined data sets, despite the significant differences in the respective study sample universes.

these data sets have considerably different sample universes both spatially and temporally. When displayed visually, the similarities are nothing short of striking. They suggest that the results seen in the Sunshine data are not a fluke, but reflect a greater pattern of prosecution and defence dynamics throughout the United States over time.

4.4 Modelling

4.4.1 Multinomial Logistic Regression

Though the above results in the Sunshine data set may be suggestive, refining the estimates for the impact of race on the probability of rejection and controlling for the myriad of possible confounding factors motivated the fitting of multiple models. This is helped by the nature of the mobile plots used in Figure 1 and elsewhere. The mobile plot displays the conditional distribution of a categorical variable with multiple levels, which suggests modelling with a conditional multinomial distribution. Consequently, fitting multinomial log-linear regression seemed a natural choice.

An equally valid choice would have been Poisson regression. As the conditional distribution of a particular outcome given some marginal count is multinomial for Poisson distributed random variables, these two models are essentially equivalent (Baker (1994); Lang (1996)). Multinomial regression was chosen for greater interpretability, as in a Poisson model the parameters of interest are all interactions as opposed to main effects. The specific implementation of multinomial regression utilized is the `multinom` function in the `nnet` package in R, which implements a method of fitting multinomial models discussed in Venables and Ripley (2002).

For all models, the pivot outcome chosen was the probability of a venire member sitting on the final jury, in other words the probability their disposition was coded as `Kept`. Coefficients were then estimated using treatment contrasts with a black female venire member with Democrat voting tendencies in a case with a black female defendant used as the reference treatment.

While the choice of race, gender, and political affiliation for the reference level was not deliberate, the choice of pivot outcome, that of a venire member being kept, was made in order to make the visualizations constructed using the coefficients clearer and easier to compare to previous visualizations. Previous visualizations, including the mobile plots, have displayed the conditional probabilities of removal with cause and strikes by defence and prosecution, and if any were used as the pivot their coefficient estimates would be hidden as the intercept.

The mobile plots created in 4.3 and 4.2 suggest, primarily, that an interaction between race and defendant race is relevant in modelling the conditional probability of venire member rejection. These plots do not suggest that any other interactions are likely to be significant. This motivated a model of all main effects and an interaction effect between the race of the venire member and that of the defendant, call it the full model. To test this interaction and the impact of venire member race in the presence of all other variables, models excluding this interaction and excluding venire member race were also fit, call them reduced models 1 and 2 respectively. The deviance residuals of all three models were used to perform ANOVA tests comparing **??** and **??** to **??**. These can be see in Table 2, which compares the deviances of the different models sequentially when fit to the Jury Sunshine Data.

Table 2: Comparison of models, displaying the residual deviance, residual degrees of freedom, differences, and p-values of these differences for adjacent models.

Model	Residual df	Residual Deviance	Difference	$P(\chi^2)$
Reduced Model 2	55527	39496		
Reduced Model 1	55521	39087	405	0
Full Model	55509	39023	67	1.4e-9

Even when controlling for defendant characteristics and the venire member's political affiliation

and sex, the race of the venire member and its interaction with the defendant race are both highly significant at the $\alpha = 0.05$ level. This suggests that the rejection of venire members is, at least in part, based on their racial characteristics. Note that the residual deviance values indicate that this model is underdispersed, implying that the significant test results gained are conservative.

Due to its significance compared to the other models, **??** was taken as the final model to estimate the race effects with precision. The estimated coefficients for the different effects and their approximate 95% confidence intervals, calculated using the standard errors of the coefficients and using a normal assumption, are displayed in Figure 9 **??**.

This table is somewhat daunting and difficult to interpret, as predicted by 4.2. No sense for the relative magnitudes of the coefficients or their significance is offered by the precise recording of all values. This motivated further visualization of these parameter estimates in an intuitive and simple way. Inspired by Solt and Hu (2018), a custom dot and whisker plot was generated. This plot displays the coefficient estimates as points in the centre of a line, the endpoints of which give the associated confidence intervals. A vertical line is placed at zero to provide a reference for the sign and significance of different parameters. Modifying this concept to suit the multinomial regression data was simple, as it only required grouping the different possible outcomes for each parameter spatially. Figure 9 displays these parameter estimates in the dot-whisker format.

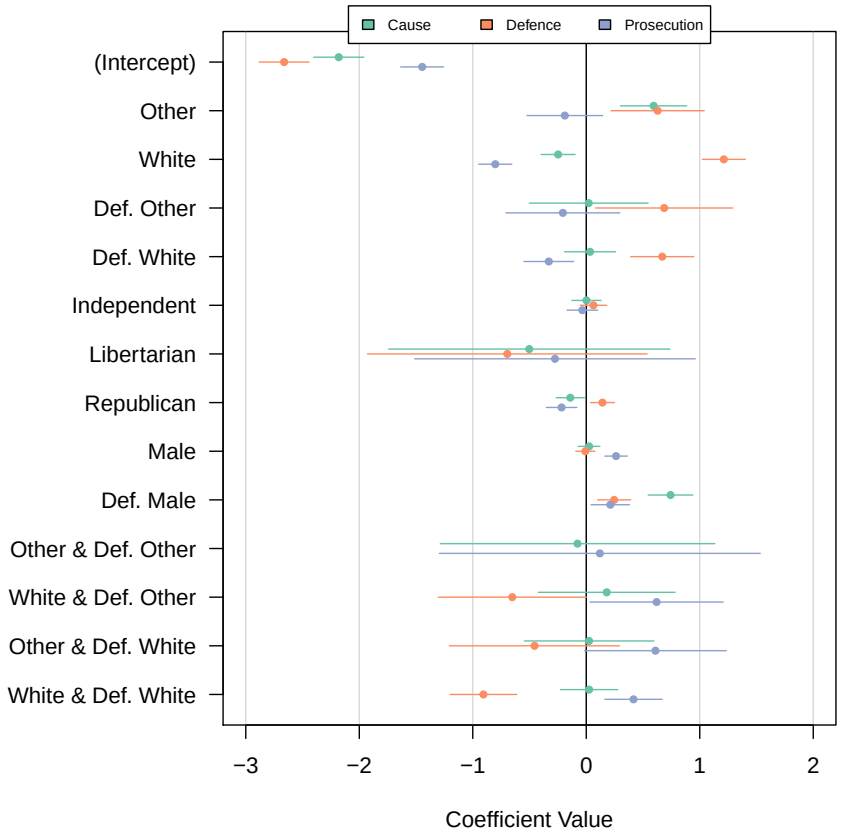


Figure 9: Model coefficients from the full model, **??**, displayed using a dot-whisker plot. The lines indicate the confidence intervals while the central points indicate the point estimates of coefficients.

This plot is considerably less daunting than the table, and provides a much clearer picture of the magnitudes, signs, and significances of the variables. However, there is still considerable visual clutter. As was noted as early as Figure 2, the sample sizes in the Sunshine data set corresponding to racial groups other than white or black are too small to make statements of significance. This same pattern is observed here, where all coefficients estimating the effects of other racial groups have wider confidence intervals and are rarely significant. Additionally, the lack of libertarian voters creates large uncertainty about the estimates of this parameter as well. To provide a clearer picture of the racial coefficient estimates and the effect of the other controls, a second dot-whisker plot was generated excluding these parameter estimates in Figure 10.

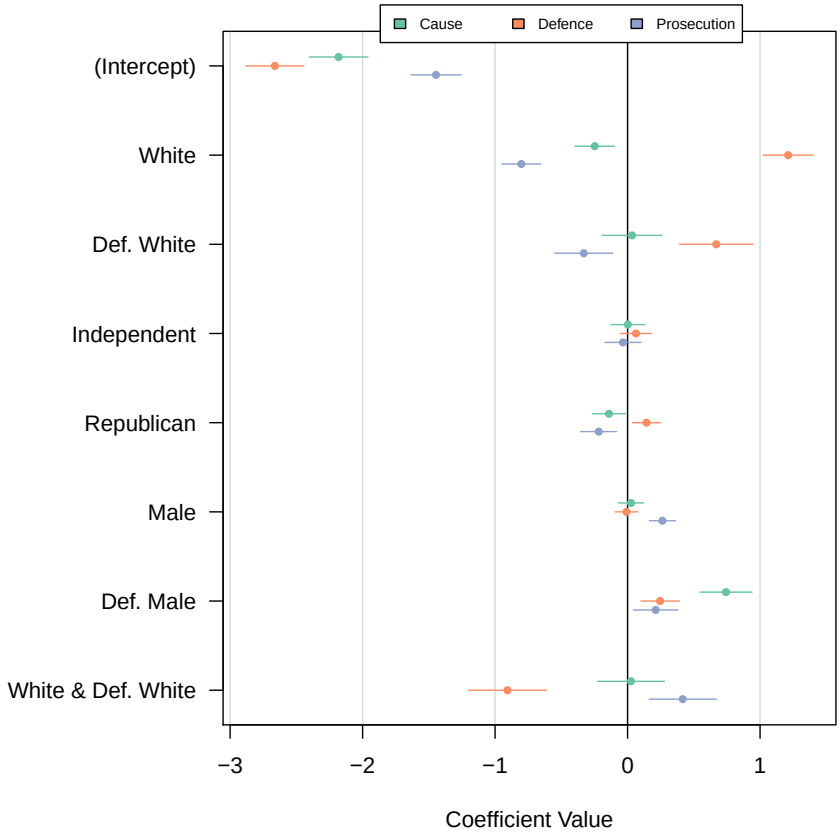


Figure 10: Select model coefficients from the full model displayed using a dot-whisker plot. The lines indicate the confidence intervals while the central points indicate the point estimates of coefficients.

The first feature of this plot to notice is the position of the coefficient estimates relative to the black line at zero. This indicates the sign of the coefficient, positive if the point is to the right and negative if it is to the left. A positive sign indicates that the factor corresponding to the coefficient increases the probability of a venire member being struck by a particular party and a negative sign indicates that the presence of said factor decreases the probability of a venire member being struck. The pattern of positive and negative values suggests the uncontrolled patterns noticed in the mobile plots were not the result of confounding.

First, take a broad overview of the patterns. The variables which show the greatest significance and the characteristic oppositional tendencies of the prosecution and defence are all race variables. The effect of a white venire member, white defendant, and their interaction all display significant, and opposite, tendencies for these two oppositional parties. Moreover, the magnitude of the defence coefficients is always greater than the prosecution for the race variables. This suggests that the defence probability to strike is affected far more by race variables than the prosecution probability to strike. To put this more simply, the defence is more sensitive to the racial aspects of the trial in general.

A quick survey of the prosecution and cause coefficients shows far less agreement than seemed apparent in the mobile plots. Scanning from the top to the bottom, the cause coefficients match the prosecution for a few groups, but are generally quite different and are often more similar to the defence than the prosecution. The suggestive pattern of matching prosecution strikes and causal strikes viewed in 4.2 vanishes when controls are placed on possible confounders[9]. In general, the cause strike coefficients are lower in magnitude than both the prosecution and defence, with the notable exception of the effect for male defendants, where it is significantly greater than both.

In order to examine patterns in more detail some important differentiation is necessary between the intercept values and the other coefficients. As with all linear models, the intercept values provide a locational measure. That is to say, the intercepts provide the base level upon which the coefficients act. Consequently, the negative values of the intercept for all striking parties is to be expected. In the Sunshine data more venire members were kept than rejected, as can be seen in Tables 1 and **??**, and so the intercept, which gives the log odds ratio of each strike outcome to the venire member being kept, would naturally be negative for all parties.

The more important feature to note for this intercept is the large difference between the different striking parties, without overlapping confidence intervals. These suggest that all three parties behave differently in the reference group, which consists of a female, black, Democrat venire member in a case with a female, black defendant. Crucially, the defence intercept is far lower than the prosecution, matching what was observed in Figure 1, where the lowest defence strike probability occurred for black venire members in cases with black defendants. Also visible in Figure 1 is an increase in defence strike probability and decrease in prosecution strike probability for a white defendant with a black or other venire member. This pattern is reflected perfectly in the positive defence coefficient and negative prosecution coefficient for white defendants.

The coefficients for white venire members also match the expectations of Figure 1. The defence attains its largest positive coefficient for this group and the prosecution is largest negative value. The dominant effect of venire member race visible throughout the mobile plots in 4.3 and 4.2 is reflected by the dominance of the magnitudes of the coefficients for white venire members relative to the reference black venire members. The defence is much more likely to reject a white venire member and the prosecution much less likely.

Slightly attenuating this pattern is the interaction effect between a white defendant and a white venire member. For this specific combination, the defence is less likely to reject than for a white venire member with a black or other defendant, while the prosecution is more likely to reject. Both of these trends seem to be significant based on the plotted confidence intervals. Combine this observation with the almost mirrored pattern for a white defendant with a black venire member (given by the white defendant coefficient), and a more nuanced strategy becomes clear. While

[9]This serves as an excellent demonstration of how complex multiple confounders can create spurious patterns.

the prosecution dominantly rejects black venire members and the defence dominantly rejects white venire members, they are also sensitive to possible racial matches. The defence pattern is consistent with a strategy which aims to partially select for venire members which have the same race as the defendant, while the prosecution has the opposite pattern.

This pattern is precisely the problematic behaviour which was implied by the individuals that viewed *R. v. Stanley* as a travesty. While the data here *cannot* reveal anything about that *specific* trial, or indeed about the exercise of peremptory challenges in Canada generally. That such a pattern is clearly present in the data will no doubt feel vindicating for those still musing on *R. v. Stanley*'s perceived injustice.

Such racial patterns in the prosecution and the defence dominate all other effects. The political effect is minor, though consistent with the hypothesis put forward by Revesz (2016), with a preference for Republicans by the prosecution and a preference for Democrats by the defence. Interestingly, all parties seem to have an increased strike probability for male venire members, suggesting a universal preference for female venire members on the jury. One would hope that this pattern is for some other reason than anything listed on pages 152-153 of Van Dyke (1977) and reproduced here in 4.3.1. While no detailed explanation is forthcoming for this pattern in the presence of racial and political affiliation controls, it would certainly be an interesting avenue for further research, as the exercise of peremptory challenges based purely on gender is also protected by *J.E.B. v. Alabama* [Supreme Court of the United States (1993)].

A lack of consistent information between data sets and of time made fitting additional models using the other data sets more difficult. Regardless of the lack of specific coefficient comparisons due to these difficulties, the above findings are consistent with the modelling results in Grosso and O'Brien (2012) and Baldus et al. (2001). Both of these studies performed logistic regression analysis of the prosecution strike patterns, controlling for attitudinal variables. In both cases, the venire member race remained a significant predictor of the venire member removal. Critically, Baldus et al. fit analogous models for the defence strike use and found the race remained significant even when the attitudinal variables were controlled. Additionally, these models found the prosecution struck black venire members at a higher rate and the defence at a lower rate, exactly as found here. The results of the modelling performed here add evidence to the mounting case of racial importance in the execution of peremptory challenges.

4.5 Trial Level Summary

While Wright et al. (2018) reported a great deal of aggregate statistics about the venire members themselves, one piece of investigation which was lacking was an analysis which viewed the trends for cases, rather than simply for individual venire members. As it cannot be known why a particular venire member is struck, and viewing their aggregate statistics tells us nothing about how different strikes relate to each other, it is possible the aggregate patterns are the result of some effect which is not due to persistent bias across trials.

By aggregating the venire members by trial and viewing the demographic trends in strikes and behaviour at this level, detailed insight into the impact of challenges at a more relevant scale is gleaned. Additionally, such aggregation allows for the synthesis of certain measures, such as a disitributional difference via the Kullback-Leibler divergence [Kullback and Leibler (1951)], which would otherwise not be well defined. This particular perspective of the data has also not been explored by any other studies known to the author.

4.5.1 Estimating Struck Juror Counts

One gap which was present in the Sunshine data was the total count of defence and prosecution removals for each trial. This variable is of minor importance for modelling the individual venire members, but it is of major interest when viewing the trials themselves. While counts of these strikes were provided in the data, there were many missing values, or values inconsistent with the number of associated venire members in the data. For example, many of the recorded values in these columns were zero even when venire members associated with that trial were marked as struck.

To remedy this, a number of variables were generated for each trial. Primarily, these variables were counts of the number of venire members with particular characteristics which were struck by both the prosecution and defence in the trial. For example the variable `DefRem.Race.Black` provided the count of black venire members struck by the defence for a specific trial. To provide an estimate of the number of venire members struck by each party, these counts were summed for one particular variable, typically gender. The sum was then compared to the recorded value for that party's removed count. The greater of the two values was kept as an estimated count to be used in analysis. For both the prosecution and the defence, about 80% of the sum and recorded values matched exactly and about 18% of the recorded values were less than the sum. This suggests some incompleteness for the remaining 2–4% of the data.

4.5.2 Visualizing the Racial Trends

The Positional Boxplot:

Motivated by the plots in the `extracat` package in R [Pilhöfer and Unwin (2013)], in particular the `rmb` plot, a positional proportion plot, called here the "positional boxplot," was developed to display the strike tendencies across trials. The data for each trial consists of categorical variables and two integer counts corresponding to the estimated defence and prosecution strike counts described in 4.5.1. The positional boxplot was developed to investigate the relationship between the strike patterns of both parties and the defendant race.

Therefore, the positional boxplot is designed to visualize the conditional distribution of a categorical variable given two integer count variables. Each observation consists of a level of a categorical variable and two counts. Across the whole data, there may be occurrences of identical values for the two counts. At each unique combination of integer counts observed, a box is placed with an area proportional to the number of observations with that specific combination. Each box is then subdivided horizontally such that the width of each subdivision is proportional to the corresponding count of each level of the categorical variable that occurs with that specific integer combination.

Figure 11 displays a positional boxplot of defendant race to prosecution and defence estimated strike counts. While the internal box patterns look rather similar everywhere, split somewhat evenly between black and white defendants, their areas are quite revealing. In the aggregate examination of the data, it was clear that the defence had an overall higher strike rate than the prosecution, this effect can be viewed clearly in Figure 7 where the defence has a higher strike rate across all gender combinations. What was not clear was whether this pattern reflected the typical pattern at the individual trial level or was rather due outlier cases with abundant defence strikes. The areas of the boxes in Figure 11 indicate clearly that the defence typically uses more peremptory challenges than the prosecution in each trial.

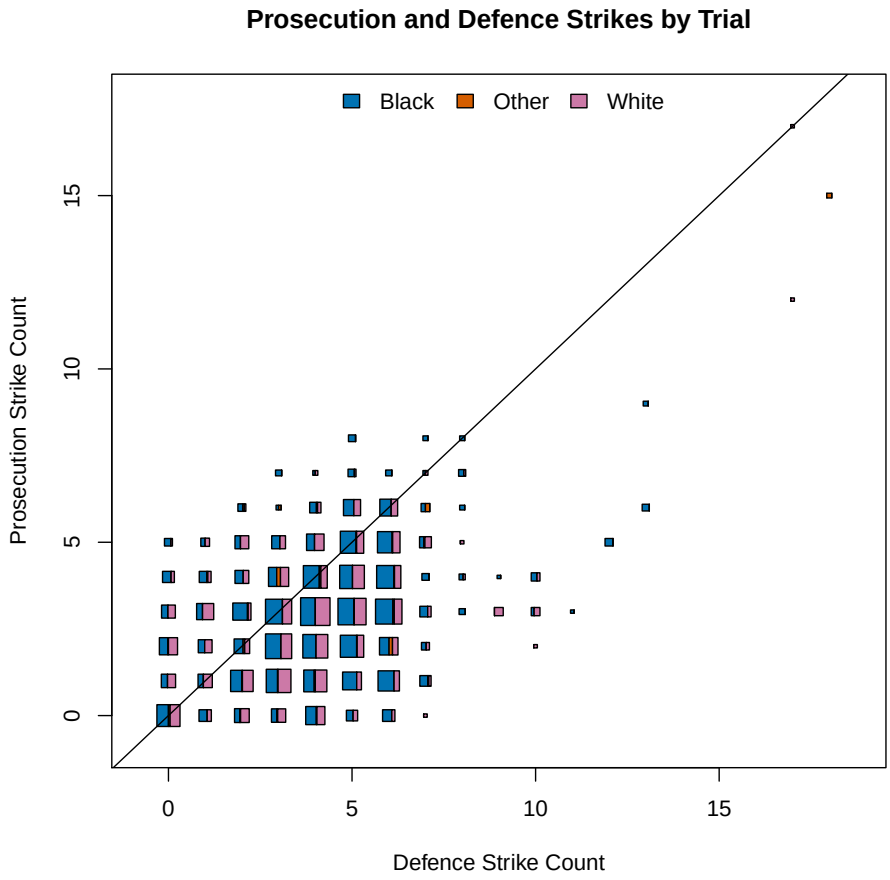
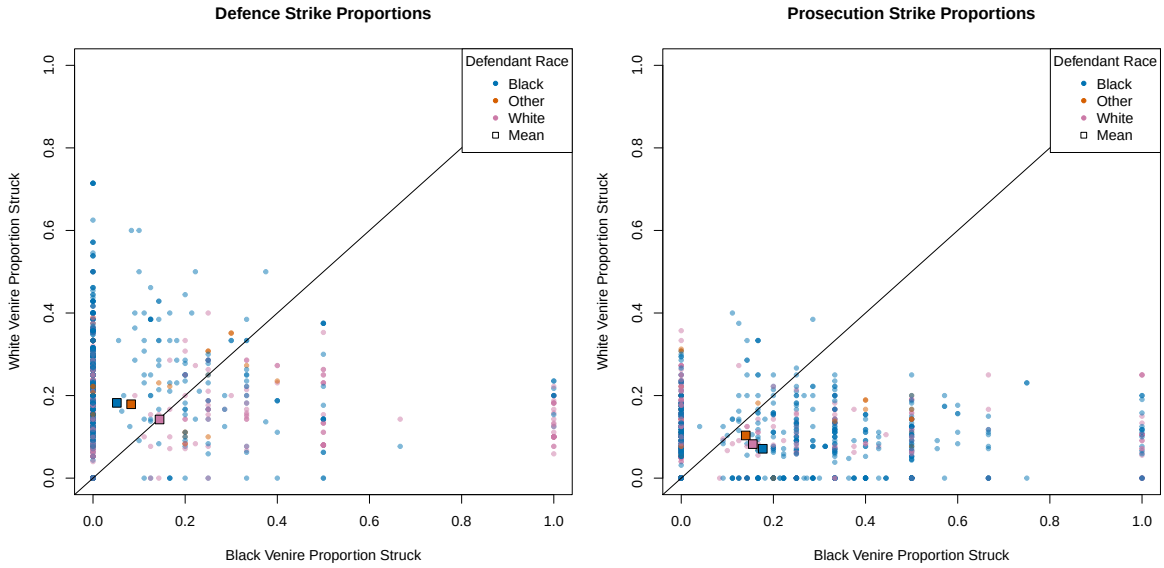


Figure 11: The positional boxplot of strikes by race of defendant for the Sunshine data. There does not seem to be a dependence of strike counts on the defendant race, the boxes look similar across the entire plot.

This plot may be somewhat misleading, however, as it fails to account for the number of black and white venire members presented to the court in each of the trials. These numbers are very different, as can be seen by the lengths of the horizontal black lines in Figure 1. A sizeable majority of the venire members are white. Luckily, with trial aggregated data, the raw strike counts can be normalized into proportions. The resulting scatterplots of these proportions by trial are displayed in Figure 12.

Here, the prosecution and defence biases are much more clear. The prosecution never strikes more than 40% of white venire members presented, and on average strikes a greater proportion of the black venire than the white venire. The defence, in contrast, regularly strikes more than 40% of the white venire, and on average strikes a greater proportion of the white venire members than the black venire members. This reinforces the aggregate observations made in 4.4, indicating that these mechanics operate on the individual trial level, and not just over all trials. Additionally, the high variation visible in Figure 12 suggests that the aggregate patterns described in 4.4 are not followed by all defence or prosecution lawyers, merely on average.

While these differences are intriguing, the far more interesting observation that can be gleaned from these plots is the fundamental difference in inclusion of minority and majority groups in jury



(a) Defence proportion venire removed

(b) Prosecution proportion venire removed

Figure 12: Scatterplot of venire proportion removed by race for both the defence and prosecution.

formation. The aggregate statistics indicate that the black venire members are a minority, and Figure 12 suggests that as a consequence, it is common that a majority or all of the potential black jurors will be removed by peremptory challenge. Such an occurrence is not observed in this data for the majority white venire members, suggesting the complete removal of white venire members is incredibly rare, if not impossible. While in many trials all black venire members were struck, in not a single trial was every white venire member struck. This suggests that if one had a strategy of keeping minorities off of juries, the peremptory challenge system would make this task easier. Such an observation is obvious in theory, but it is no less impactful to see it emerge from the data.

4.6 On Venire Selection

Van Dyke (1977) spends much of chapters four and five exploring the causes for the under-representation of certain groups in jury venires, and his analysis suggests that under-representation starts at the jury selection stage. Fewer non-white individuals register to vote generally (page 89). Additionally, excusal from jury duty, primarily granted for economic hardship, impacts disadvantaged economic groups to a greater extent (pages 113-120).

Such explanations provide a plausible reason why black males would be most under-represented in venires, and are also possibly contributing factors to the white majority of the venire despite the majority of defendants being black. Such issues with the jury selection process will not, and cannot, be solved by simply removing the peremptory challenge. They have much more to do with the relationship between minority groups and wider society, and so require more comprehensive and complex solutions.

R. v. Stanley serves as an excellent demonstration of this problem. As Quenneville and Warick (2018a) report, 750 individuals on the jury roll were sent court summons, but only 204 of these individuals appeared on the day of the trial. This large non-response population introduces the

potential for incredible bias in the venire. If there is some relationship between demographics and an inability to serve on a jury or a lack of will to respond to a court summons, as Van Dyke (1977) suggests with ample evidence, then problems with minority representaton may be due in larger part to the method of jury summons than the exercise of challenges in the courtroom. Removing the peremptory challenge may serve as a useful step in the right direction, but it cannot be expected to "fix" minority representation alone.

5 Summary

The visual tools and models presented here support the dominant analysis in the literature. The primary determinant in the exercise of peremptory challenges is race in the Sunshine data set. The prosecution tends to remove more racial minority venire members than expected and fewer white venire members than expected. The defence tends to have the opposite strategy. This pattern is not only seen in aggregate in 4.2, but is visible in the trial summaries presented in 4.5. The impact of race remains apparent in the mobile plots even when other legitimate factors such as political affiliation are controlled.

Beyond detecting the patterns in one data set, this work demonstrates other strengths of the mobile plot and visual analysis. The first of these is the utility of the mobile plot to compare the strike patterns across multiple data sets. The similarities between the Stubborn, Philadelphia, and Sunshine data sets are immediately clear when visualized appropriately. This is critical in examining the practice of peremptory challenges, as it allows for a comparison of their use across studies with radically different study populations so long as analogous data is collected.

In this case, the strength of the similarities observed between these data sets when visualized with the mobile plot suggests the pattern of racial preferences is not a local phenomenon in location or time, but is a reflection of a strategy utilized by the prosecution and defence in jury trials generally. As more data and investigation take place, further visual comparisons can motivate more analysis of the similarities between these patterns in different studies. Based on the findings here and in the other empirical analyses, the critical data which should be collected for analysis is the disposition of the venire members and the demographic characteristics of both the venire member and defendant. Attitudinal data, which was not standardized in the data analysed in this work, should also be collected and analyzed in some standard way to augment the utility of the mobile plot for comparison.

Another strength of the mobile plot is the motivation of model building. The multinomial regression models from 4.4 were created to fit models analogous to the displayed patterns of the mobile plots generated in advance. That the findings of the models matched the analysis of the mobile plots almost exactly justifies these plots as informative analytical tools. The models allowed for the estimation of effects in the Sunshine data controlling for possible legitimate confounders, giving a table of coefficient estimates consistent with those generated previously for other data sets, such as the Stubborn and Philadelphia data.

It was not until these coefficients were visualized with the dot-whisker plot, however, that a number of more nuanced patterns became obvious. The first of these is the greater sensitivity of the defence to the racial aspects of a trial than the prosecution. That is, the race of the venire member has a greater impact on the defence's probability of rejection than the prosecution's. The second pattern is the tendency of race matching by the defence and race contrasting by the prosecution.

This aggregate pattern also seems to be reflected in the trial level summary of the data, which suggests that this trend is not a quirk of aggregation, but a reflection of individual lawyer decision making.

Of course, as suggestive as these patterns are, and as spotted as the history of peremptory challenges is with controversy, none of this can say definitively whether the individuals rejected from the Sunshine venires were rejected appropriately due to their "extreme" bias. Without detailed descriptions of the bias of the population as a whole, such judgements on the propriety of strikes simply cannot be made, and whether these racial strike patterns are simply the result of legitimate strikes issued for reasons related strongly to race remains unknown.

Despite this limitation, the final scatterplots suggest a criticism of peremptory challenges independent of these concerns which is consistent with the source of controversy for *Batson v. Kentucky*, *Swain v. Alabama*, and *R. v. Stanley*. Peremptory challenges frequently remove all representatives of minority groups from the venire. This prevents their participation in a panel meant to represent the conscience of their community, corroding a critical function of the jury and creating a group sceptical of the operation of the legal system. Certainly, the smaller sizes of minority groups and their relationship to the majority may lead to their under-representation for reasons other than peremptory challenges, but a graphical exploration shows definitively that a component of their under-representation is peremptory challenges. Minority groups are fully struck from the venire by peremptories far more often than majority groups. Striking all majority members may, in fact, be virtually impossible.

All of this paints a bleak picture of the role of peremptory challenges in the modern jury selection process. Without additional work, it is impossible to say with certainty whether the racial patterns observed here and elsewhere are due to racial prejudice by the court, but one may ask the question of whether that matters. As Lord Chief Justice Hewart said in *R. v. Sussex Justices*

> Justice should not only be done, but should manifestly and undoubtedly be seen to be done

The visualizations of this paper have made much seen, but it is doubtful that what has been seen here, and by the critics of the *R. v. Stanley* proceedings, looks like justice.

6 Future Work

One obvious way to extend the work done here is through more thorough modelling. While the multinomial regression model fit in 4.4 served its purpose, much more precise models could be fit using causal graphs. Such causal modelling has the possibility to extend the observations of the model from the simple pattern identification of the multinomial model and visualizations presented here to precise statements about the magnitude of causal effects between factors. Representing the factors in a causal graph would also be a useful exercise in making the assumptions of the model abundantly clear. Logistic regression models and multinomial models, which have been the norm for peremptory challenge data so far, are less clear about their assumptions, especially to those not trained to fit and analyse these models.

Other possible models of interest are mixed models. In this work the attempts to fit mixed models were not discussed, but at several points models with random effects for each trial were attempted. Unfortunately, these models failed to converge. Not a great deal of time was spent

trying to transform the data to facilitate convergence to a reasonable value, and so no mixed models were fit. Such models are attractive because they have the potential to flexibly control for a host of factors which will vary between trials, and do so in a manner which involves minimal parameters. Estimating a random effect for lawyer, for instance, could shed light on how variable lawyers are in their behaviour. This dimension of individual variability is essentially unadressed by the aggregate examinations of this work.

Another extension would be further investigation of the Sunshine data. It is an incredibly rich data set and this work only examined one small facet of it. The crime classification outlined in 3.1, for example, was never utilized in the analysis, despite the investment of time and effort in performing this clean up. Perhaps this method could also be applied to other irregular data in court cases or elsewhere to efficiently categorize irregular strings.

Finally, as Wright et al. (2018) notes, more data is needed on this topic generally. Further efforts to collect data and reinforce or refute the findings of this work and previous ones should be undertaken, and efforts to centralize and regularize the data would assist in the ease of analysis. Using the visualizations of this paper, such as the mobile plot and positional boxplot, quick and informative comparisons of new data to older data sets could be performed easily. Increased transparency and centralized data collection additionally have the potential to allow for a greater understanding of which elements of the jury trial system work and which are inappropriate. As Wright et al. puts it:

> The transformative power of data ... could improve the courts' efforts to regulate the work of other criminal justice players, such as police and prosecutors.

References

42nd Parliament of Canada (2018a). Bill C-75: An Act to Amend the Criminal Code, Youth Criminal Justice Act and other Acts and to make consequential amendments to other Acts. http://www.justice.gc.ca/eng/csj-sjc/pl/charter-charte/c75.html (Accessed February 2019).

42nd Parliament of Canada (2018b). Bill C75. LEGISinfo. http://www.parl.ca/LegisInfo (Accessed February 2019).

Attorney General's Office of the United Kingdom (2012). Jury vetting right of stand-by guidelines. https://www.gov.uk/guidance/jury-vetting-right-of-stand-by-guidelines–2 (Accessed February 2019).

Baker, S. G. (1994). The multinomial-poisson transformation. *Journal of the Royal Statistical Society, Series D (The Statistician)*, 43(4):495–504.

Baldus, D. C., Grosso, C. M., Dunham, R., Woodworth, G., and Newell, R. (2012). Statistical proof of racial discrimination in the use of peremptory challenges: The impact and promise of the miller-el line of cases as reflected in the experience of one philadelphia capital case. *Iowa Law Review*, 97.

Baldus, D. C., Woodworth, G., Zuckerman, D., and Weiner, N. A. (2001). The Use of Peremptory Challenges in Capital Murder Trials: A Legal and Empirical Analysis. *University of Pennsylvania Journal of Consitutional Law*, 3(1).

Broderick, R. J. (1992). Why the peremptory challenge should be abolished. *Temple Law Review*, 65(2):369.

Brown, F. L., McGuire, F. T., and Winters, M. S. (1978). The peremptory challenge as a manipulative device in criminal trials: Traditional use or abuse. *New England Law Review*, 14:192.

Cleveland, W. S. and McGill, R. (1987). Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society. Series A (General)*, 150(3):192–229.

Flanagan, F. X. (2015). Peremptory Challenges and Jury Selection. *The Journal of Law and Economics*, 58.

Ford, R. (2010). Modeling the effects of peremptory challenges on jury selection and jury verdicts. *George Mason Law Review*, 17:377.

Government of Canada (1985). Criminal code section 634. https://laws-lois.justice.gc.ca/eng/acts/C-46/section-634.html (Accessed February 2019).

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

Hans, V. P. and Vidmar, N. (1986). *Judging the Jury*. Plenum Press, 1 edition.

Harris, K. (2018). Liberals review jury selection process after Boushie case uproar. CBC News. https://www.cbc.ca/news/politics/jury-selection-diversity-indigenous-1.4531792 (Accessed February 2019).

Hasan, N. R. (2018). Eliminating peremptory challenges makes trials less fair. The Star. https://www.thestar.com/opinion/contributors/2018/04/10/eliminating-peremptory-challenges-make-trials-less-fair.html (Accessed February 2019).

Hoffman, M. B. (1997). Peremptory Challenges Should Be Abolished: A Trial Judge's Perspective. *The University of Chicago Law Review*, 64(3):809.

Iacobucci, F. (2013). First Nations Representation on Ontario Juries: Report of the Independent Review Conducted by the Honourable Frank Iacobucci. Ministry of the Attorney General. https://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/ (Accessed February 2019).

Kullback, S. and Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

Lang, J. B. (1996). On the comparison of multinomial and poisson log-linear models. *Journal of the Royal Statistical Society, Series B (Methodological)*, 58(1):253–266.

MacLean, C. (2018). Gerald Stanley acquittal renews calls for justice reform 27 years after Manitoba inquiry. CBC News. https://www.cbc.ca/news/canada/manitoba/aboriginal-justice-inquiry-colten-boushie-gerald-stanley-jury-1.4532394 (Accessed February 2019).

Macnab, A. (2018). Stanley acquittal should not lead to scrapping peremptory challenges, say criminal lawyers. Canadian Lawyer. https://www.canadianlawyermag.com/legalfeeds/author/aidan-macnab/stanley-acquittal-should-not-lead-to-scrapping-peremptory-challenges-say-criminal-lawyers-15332/ (Accessed February 2019).

Morehead, J. W. (1994). When a peremptory challenge is no longer peremptory: Batson's unfortunate failure to eradicate invidious discrimination from jury selection. *DePaul Law Review*, 43:625.

Nunn, K. B. (1993). Rights held hostage: Race, ideology, and the peremptory challenge. *Harvard Civil Rights-Civil Liberties Law Review*, 28:63.

Page, A. (2005). Batson's blind spot: Unconscious stereotyping and the peremptory challenge. *Boston University Law Review*, 85:155.

Pilhöfer, A. and Unwin, A. (2013). New approaches in visualization of categorical data: R package extracat. *Journal of Statistical Software*, 53(7):1–25.

Quenneville, G. and Warick, J. (2018a). 'Deck is stacked against us,'says family of Colten Boushie after jury chosen for Gerald Stanley trial. CBC News. https://www.cbc.ca/news/canada/saskatoon/colten-boushie-shooting-case-jury-selection-gerald-stanley-1.4506931 (Accessed February 2019).

Quenneville, G. and Warick, J. (2018b). Shouts of 'murderer' in courtroom after Gerald Stanley acquitted in Colten Boushie shooting. CBC News. https://www.cbc.ca/news/canada/saskatoon/gerald-stanley-colten-boushie-verdict-1.4526313 (Accessed February 2019).

Revesz, J. (2016). Ideological imbalance and the peremptory challenge. *Yale Law Journal*, 125.

Richardson Oakes, A. and Davies, H. (2016). Justice must be seen to be done: a contextual reappraisal. *Adelaide Law Review*, 37(2):461–494.

Roach, K. (2018). Ending peremptory challenges in jury selection is a good first step. The Ottawa Citizen. https://ottawacitizen.com/opinion/columnists/roach-ending-peremptory-challenges-in-jury-selection-is-a-good-first-step (Accessed February 2019).

Sison, C. P. and Glaz, J. (1995). Simultaneous confidence intervals and sample size determination for multinomial proportions. *Journal of the American Statistical Association*, 90(429):366–369.

Solt, F. and Hu, Y. (2018). *dotwhisker: Dot-and-Whisker Plots of Regression Results*. R package version 0.5.0.

Supreme Court of Canada (1991). R. v. sherratt. Supreme Court Judgements. SCC Case Number: 21501; https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/734/index.do?q=21501 (Accessed February 2019).

Supreme Court of the United States (1965). Swain v. Alabama. https://supreme.justia.com/cases/federal/us/380/202/ (Accessed February 2019).

Supreme Court of the United States (1986). Batson v. Kentucky. https://www.law.cornell.edu/supremecourt/text/476/79 (Accessed February 2019).

Supreme Court of the United States (1993). J.E.B. v. Alabama. https://supreme.justia.com/cases/federal/us/511/127/ (Accessed February 2019).

Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press, 2 edition.

Van Dyke, J. M. (1977). *Jury Selection Procedures: Our Uncertain Commitment to Representative Panels*. Ballinger Publishing, 1 edition.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.

von Moschzisker, R. (1921). The historic origin of trial by jury. *University of Pennsylvania Law Review*, 70(1).

Woolley, A. (2018). An Ethical Jury? Reflections on the Acquittal of Gerald Stanley for the Murder/Manslaughter of Colten Boushie. *Slaw: Canada's online legal magazine*. http://www.slaw.ca/2018/02/20/an-ethical-jury-reflections-on-the-acquittal-of-gerald-stanley-for-the-murder-manslaughter-of-colten-boushie/ (Accessed February 2019).

Wright, R. F., Chavis, K., and Parks, G. S. (2018). The Jury Sunshine Project: Jury Selection Data as a Political Issue. *University of Illinois Law Review*, 2018(4):1407.

Zinchuk, B. (2018). Both sides wrong about Stanley trial. Prince George Citizen. https://www.princegeorgecitizen.com/opinion/editorial/both-sides-wrong-about-stanley-trial-1.23199321 (Accessed February 2019).