Seen to Be Done: A Graphical Analysis of Peremptory Challenge

Chris Salahub

January 8, 2020

Abstract

The legal practice of peremptory challenges is described, outlining its past and present racial controversies as well as the modern defence provided in its favour. Three different peremptory challenge data sets [Wright et al. (2018); Grosso and O'Brien (2012); Baldus et al. (2001)] are analyzed for racial bias using novel visual tools including the mobile plot and the positional boxplot. Mulitnomial regression models motivated by these visualizations are fit and used to generate precise parameter estimates indicating 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 challenges in *R. v. Stanley* [Quenneville and Warick (2018)] have culminated in the abolition of the practice in Canada by Bill C-75 [42nd Parliament of Canada (2018)]. This leaves the jury selection process in Canada on somewhat shaky ground [Fine (2019)] and warrants a greater discussion of the practice than its use in one isolated case. 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. Unfortunately, there is a lack of relevant Canadian data, and so this paper aims to provide analysis and evidence to illuminate the topic further by considering three peremptory challenge data sets collected in the United States: 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 empirical 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 mathematical 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 these authors arrived 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 jury trial 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 previous analyses certainly contain all the information necessary to evaluate strike patterns, they fail to be accessible to a casual reader, as they require commitment and sharp focus to interpret and compare. In contrast, visual representations of the data could be used for quick comparison and interpretation of results and would facilitate the dissemination of empirical results of these analyses to a broader audience. This would make the work of comparing and interpreting data sets far more intuitive than the table representations which currently dominate. This work endeavours to provide such visual tools.

Consequently, this article proceeds in four parts. Section 2 provides the necessary legal context to understand the motivation of the previous investigations, outlining the general jury selection procedure and modern controversies with peremptory challenges. With the necessary context provided, Section 3 proceeds to discuss the three data sets obtained, in particular explaining the differences in data collection methods and recorded variables. Section 4 provides the details and results of the analysis performed on the different data sets. The Sunshine data serves as the centrepiece of this analysis, as it is the most comprehensive and recent of the data sets. Mobile plots are described and used to analyse the data visually, in particular to compare the Sunshine data set to the Stubborn and Philadelphia data sets. The implication of their similarities for generalization are discussed. These visual analyses motivate model selection in 4.2, garnering precise estimates of 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 onto the jury, and so becomes a juror

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

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, and this system is used in all data sets analysed here. 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 trial systems has generated a great deal of controversy. The privileged removal of a venire member without any justification has seen persistent allegations of abuse, often around the use of these challenges by prosecutors representing the state.

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. 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 since been extended to allow Batson Challenges for the sex of venire members [Supreme Court of the United States (1993)].

Racial controversies have also been present in Canada before *R. v. Stanley*. Abolition of the peremptory challenge due to alleged racial bias against First Nations venire members was recommended by the 1999 Manitoba Aboriginal Justice Implementation Commission [Aboriginal Justice Implementation Commission (1999)]. This sentiment was echoed in the Iacobucci Report on First Nations representation in Ontario juries, which recommended "an amendment to the Criminal Code that would prevent the use of peremptory challenges to discriminate against First Nations people serving on juries." [Iacobucci (2013)]

Despite these recommendations and the legal changes in the United States, 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 ... 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*: "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 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.

2.3   Previous Invesigations

The debate around peremptory challenges has already stoked academic investigation of the practice. 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 mathematical explorations by Ford (2010) and Flanagan (2015) which make use of game theory. Both investigations indicate that the current system of peremptory challenges may produce juries which are, counterintuitively, biased towards conviction or acquittal. This implies that peremptory challenges are more useful for the purpose of "stacking" a jury in favour of one side. An informative example of this practice is presented in Chapter 6 of Hans and Vidmar (1986), where 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*.

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 sole 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), and Baldus et al. (2012). 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 the 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 incorporating 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.

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 were 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 for fewer cases. This study, also based in North Carolina, detailed the trials of inmates on death row as of July 1, 2010, yielding a total of 173 cases. As in the Jury Sunshine study, variables recorded for each case are 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.

A 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 analogously between them.

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. 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

Note that the code used to perform the analysis and generate all figures is publicly available on GitHub: github.com/Salahub/peremptory_challenges.

4.1   Visualization

First, consider a simple aggregate summary of the probability of removal by peremptory challenge by race in all data sets, shown in Table 1.

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 only a weak racial bias at the aggregate level.

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

Table 1: The conditional probability of a venire member being struck peremptorily by venire member race across data sets. 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	

patterns in the data. Consequently, the "mobile plot" for visualizing 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 Figure 1 to compare dispositions of venire members by their race and the defendant race for the Sunshine data.
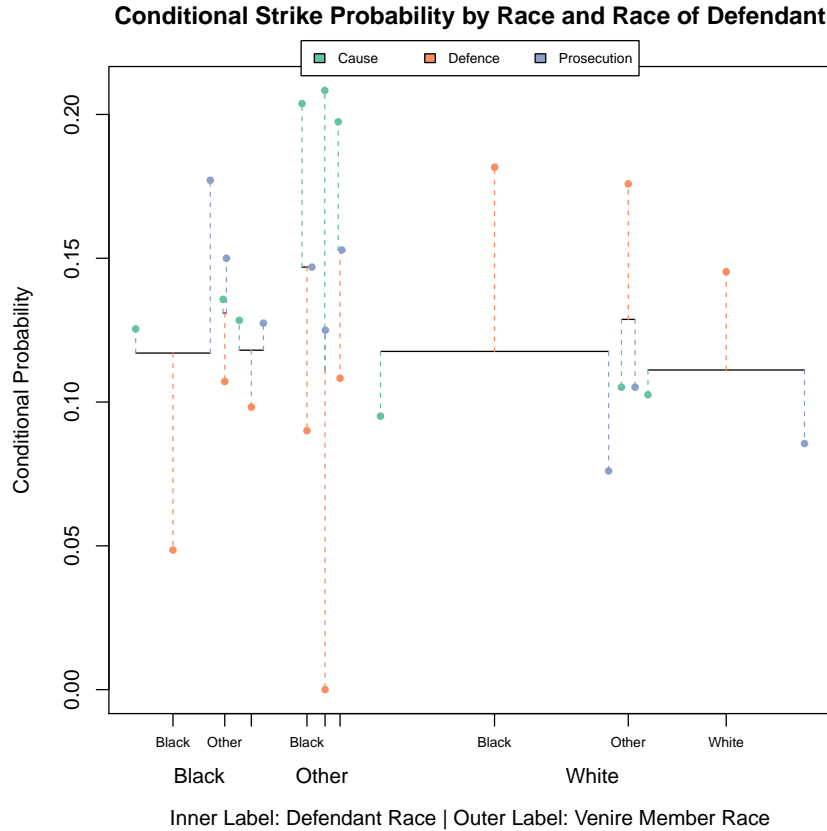

Figure 1: The conditional probability of strike disposition given the venire member and defendant race for the Sunshine data set. Expected values are represented by the horizontal black lines, and the observed values by the points at the end of the dotted lines. Each horizontal black line corresponds 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 indicated party for the given racial combination.

First, a small explanation of the 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 racial combination. 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:

$$D|D \neq \text{Kept}, R, E \sim Unif(\{\text{Prosecution Struck, Defence Struck, Struck with Cause}\}) \quad (1)$$

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 strategies. The defence seems biased towards a jury composed of more 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 a defence strike. 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. Challenges 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 afore-mentioned tendencies of the prosecution and defence are strongest for black defendants, which have the greatest departures of the conditional probabilities from the expectation. The defence 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 challenges with cause follow patterns 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. Section 4.2 discusses the agreement of these two strike tendencies further.

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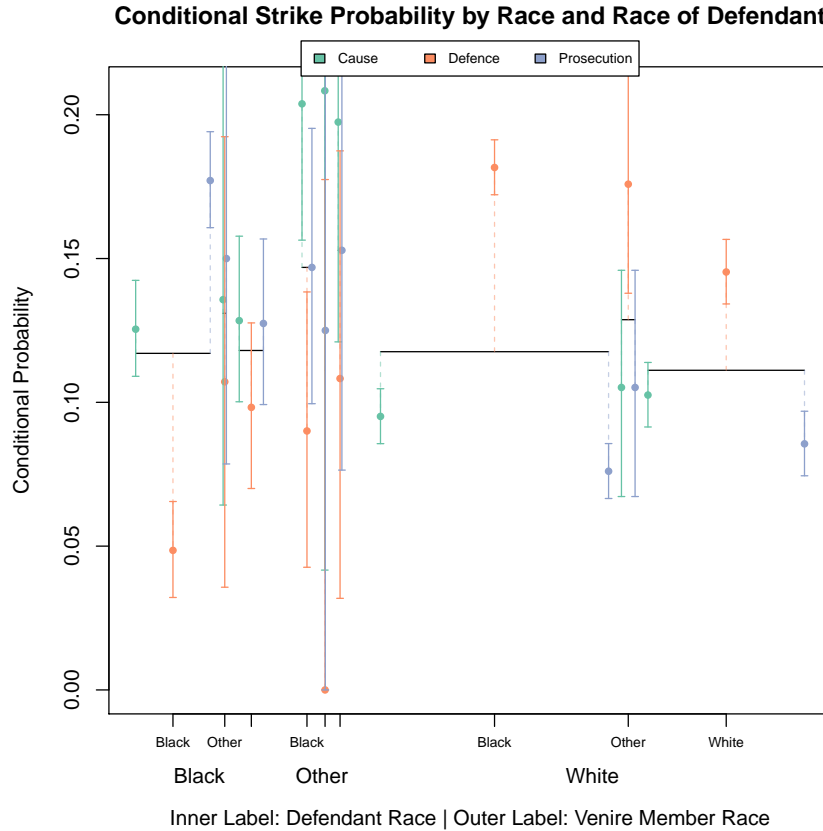
Figure 2: The plot of conditional strike probability by racial combination from Figure 1 with confidence intervals added. Note that many of the seemingly striking departures seen are insignificant when these confidence intervals are applied, especially for races other than black and white.

treme tendencies are simply a result of the well-known greater variation of samples with small sizes. In order to see the true nature of the noted departures some incorporation of the expected variation 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. However, the results for the larger groups, in particular for white venire members or black defendants, are significant. 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 Section 4.2.

4.1.1 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

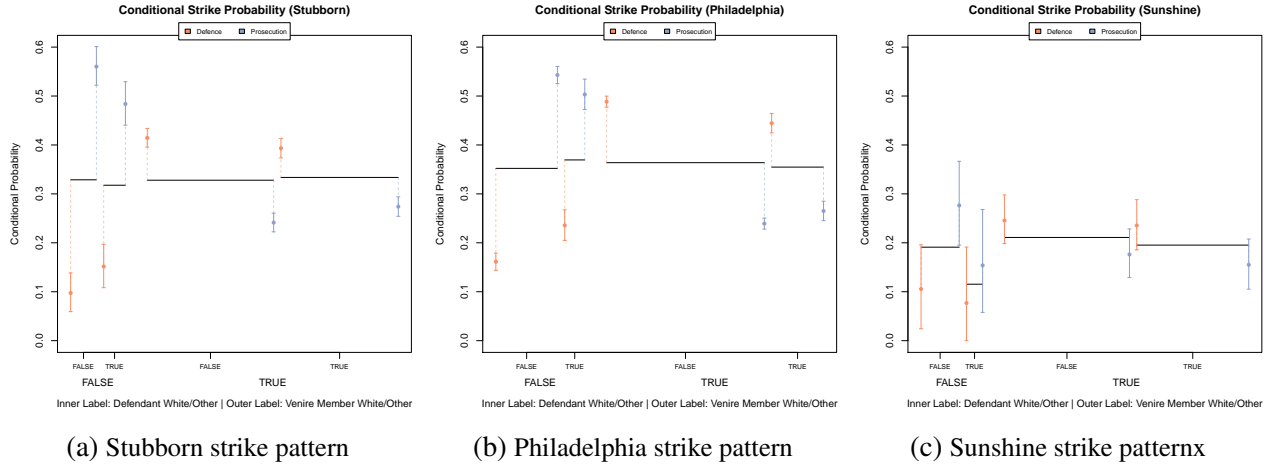| (a) Stubborn strike pattern | (b) Philadelphia strike pattern | (c) Sunshine strike patternx |

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.

comes with the ability to quickly compare different data sets. Consider, for example, comparing all three data sets 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 as the table grows in complexity. Compare this with a cursory glance at Figure 3, which makes the similarities of the data sets immediately clear. Note that the restricted study sample universes of the Philadelphia and Stubborn data make these plots considerably different than Figure 2.

Despite the very different study populations of these three data sets, all display similar patterns, with only the magnitudes of the strike rates differing. The mobile plot format reveals several interesting aspects. 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 and consistently different strategies of the defence and prosecution 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 are 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 sampling method. While the Philadelphia data and Stubborn data both collected data which included multiple years and selected capital cases only, the Sunshine data was restricted to trials which occurred in 2011 and collected all cases. 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 the Sunshine data, which is also visible in Table 1. This departure may be of interest for future investigation.

4.2 Modelling

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 *Swain v. Alabama* [Supreme Court of the United States (1965)].

Without detailed transcripts detailing voire dire, 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 related to race. Striking a venire member due to their political affiliation, for example, would be allowed, and so racial heterogeneity could cause valid strikes to appear invalid. Such an "ideological imbalance" in racial political affiliations is actually present in this data set, and has been noted elsewhere in the literature [Revesz (2016)].

Consequently, answering the question of peremptory strike validity precisely requires that all available confounding factors be controlled. While such controlled modelling has already been completed for Grosso and O'Brien (2012) and Baldus et al. (2001), no such modelling has been completed for the much richer and larger Jury Sunshine data set. The design of the mobile plot further motivates such multivariate regression by suggesting a particular form of model: the multinomial logistic regression model.

4.2.1 Multinomial Logistic Regression

The mobile plot used in Figure 1 displays the conditional distribution of a categorical variable with multiple levels given some combination of other variables. The natural model to complement this plot, then, would be a conditional multinomial model. Consequently, a multinomial log-linear regression model, or equivalently a Poisson regression [Baker (1994); Lang (1996)], was chosen to model the data. The specific implementation of multinomial regression utilized is the `multinom` function in the `nnet` package in R, which implements a method of fitting multinomial logistic regression 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 of model coefficients easily compared to previous visualizations. Choosing another pivot would have hidden its effect as the intercept, making comparison more complicated.

The mobile plots created in 4.1 suggest two main features to investigate. The first is the obvious impact of race in peremptory challenge use, and the second is the interaction between race and defendant race, as there were some slight differences seen for different racial combinations. A model of all main effects and an interaction effect between the race of the venire member and that of the defendant, called the full model, was therefore fit. To test the interaction and race effects in the presence of all other variables, nested models excluding this interaction and excluding venire member race were also fit, called reduced models 1 and 2 respectively. The deviance residuals of all three models were used to perform ANOVA tests comparing the models. These can be see in Table 2, which compares the deviances of the different models sequentially when fit to the 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, the full model was taken as the final model to estimate the race effects with precision. Inspired by the dot-whisker plot [Solt and Hu (2018)], a custom dot-whisker plot was designed to display the model coefficients. 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. 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 using this modified dot-whisker plot in Figure 4.

The first noteworthy feature of this plot is the position of the coefficient estimates relative to the black line at zero. These indicate the sign of a 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 racial patterns noticed in the mobile plots were not the result of confounding with any recorded variables.

First, view the patterns broadly. The variables which show the greatest significance and the characteristic oppositional tendencies of the prosecution and defence seen in Figure 1 are all race variables. White venire members, white defendants, and the interaction of venire member and defendant race all have significant impacts on the probability of a venire member's removal by peremptory challenge. Furthermore, the prosecution and defence display impact opposite tendencies over these racial aspects and the magnitude of the defence coefficients is consistently greater than the prosecution for the race variables. This suggests that the defence is more sensitive to the racial aspects of the trial than the prosecution on average.

A quick survey of the prosecution peremptory challenge and challenge with 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 different and often more similar to the defence than the prosecution. The suggestive pattern of matching prosecution strikes and causal strikes viewed in Figure 1 vanishes when controls are placed on possible confounders. 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
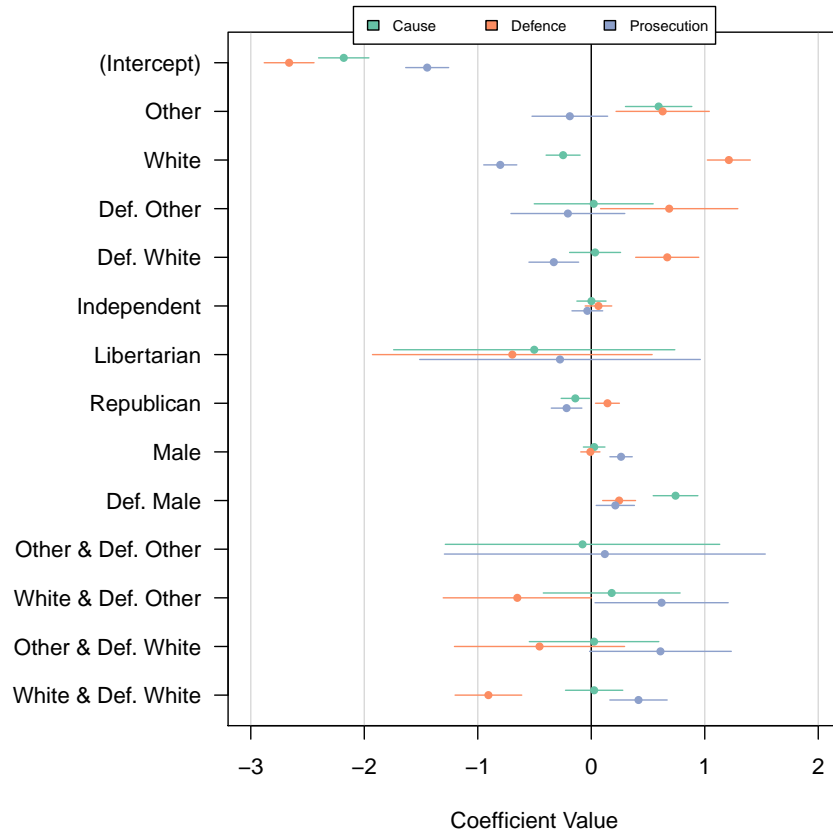
Figure 4: Model coefficients from the full model displayed using a dot-whisker plot. The lines indicate the confidence intervals while the central points indicate the point estimates of coefficients.

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. In the Sunshine data more venire members were kept than rejected, as can be seen in Table 1, and so the intercept, which gives the log odds ratio of each strike outcome to the venire member being kept, is negative for all parties as expected. The more important feature to note for the intercept is the large differences between the different striking parties. These suggest that all three parties behave differently in the pivot case, that 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 its largest negative coeffi-

13

cient, suggesting white venire members in cases with black defendants are the most polarizing. The dominant effect of venire member race visible in Figure 1 is reflected by the dominance of the magnitudes of the coefficients for white 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 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.

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.

4.3   Trial Level Summary

The control introduced by linear modelling ignores one critically important dependence that is present between jurors: the case. Cases share lawyers, judges, charges, and other important features which are not adequately controlled by the models fit in Section 4.2. As it cannot be known why a particular venire member is struck, and modelling the patterns at the venire member level does not account for the close relationship between venire members on the same case, it is possible the aggregate patterns are the result of some effect other than persistent bias across trials. By aggregating the venire members by trial and viewing the racial strike behaviour at this level, insight into the impact of challenges at a more relevant scale is gleaned.

4.3.1   Estimating Struck Juror Counts

Aggregation required some synthesis, however. 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, counts of struck venire members with particular characteristics, for example race, were computed for each trial. To provide an estimate of the total number of venire members struck by each party, these counts were summed for one particular variable, typically gender. This 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% of the data.

4.3.2 Visualizing the Racial Trends

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 Section 4.3.1. The positional boxplot was developed to investigate the relationship between the strike patterns of both parties and the defendant race.
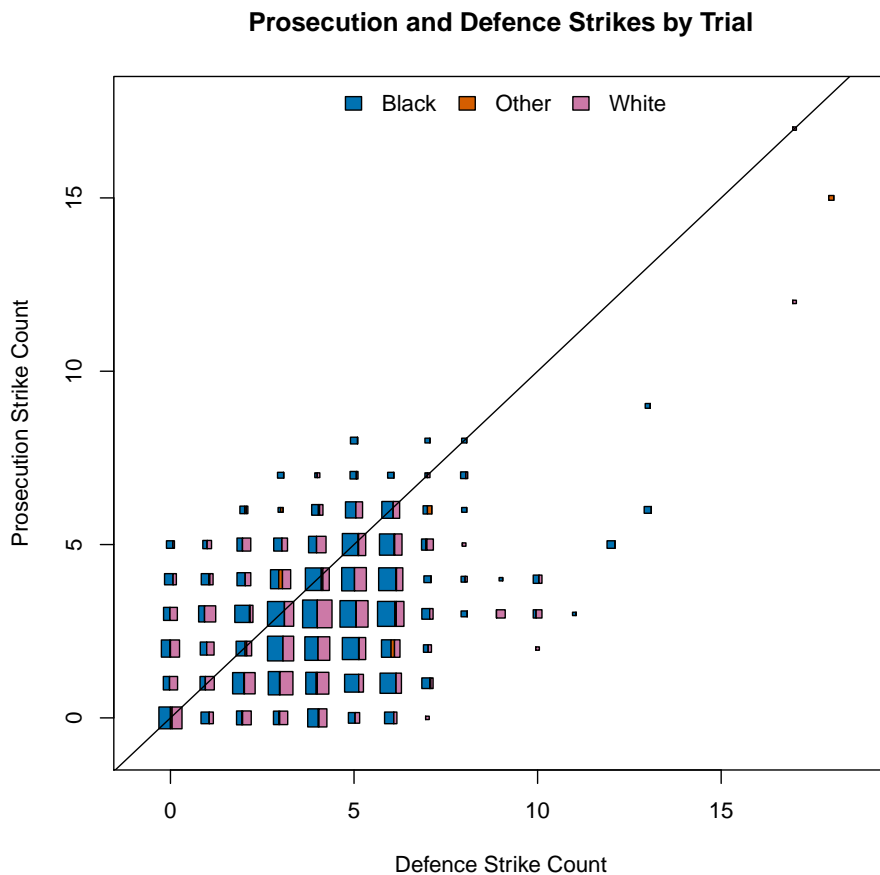

Figure 5: The positional boxplot of strikes by race of defendant for the Sunshine data. There does not seem to be a dependence of strike counts on the defendant race, the boxes look similar across the entire plot.

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

15

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 5 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. The greater area of boxes below the diagonal line demonstrates clearly the tendency for the defence to use more peremptory challenges than the prosecution in a given case.

However, this plot fails to account for the total number of black and white venire members presented to the court in each of the trials. We would expect these numbers to be very different in the Sunshine data; consider the different lengths of the horizontal black lines in Figure 1. These lengths indicate that a sizeable majority of the venire members are white. To account for this, the raw strike counts were normalized into proportions of the venire by race. The resulting scatterplots of these proportions by trial are displayed in Figure 6.

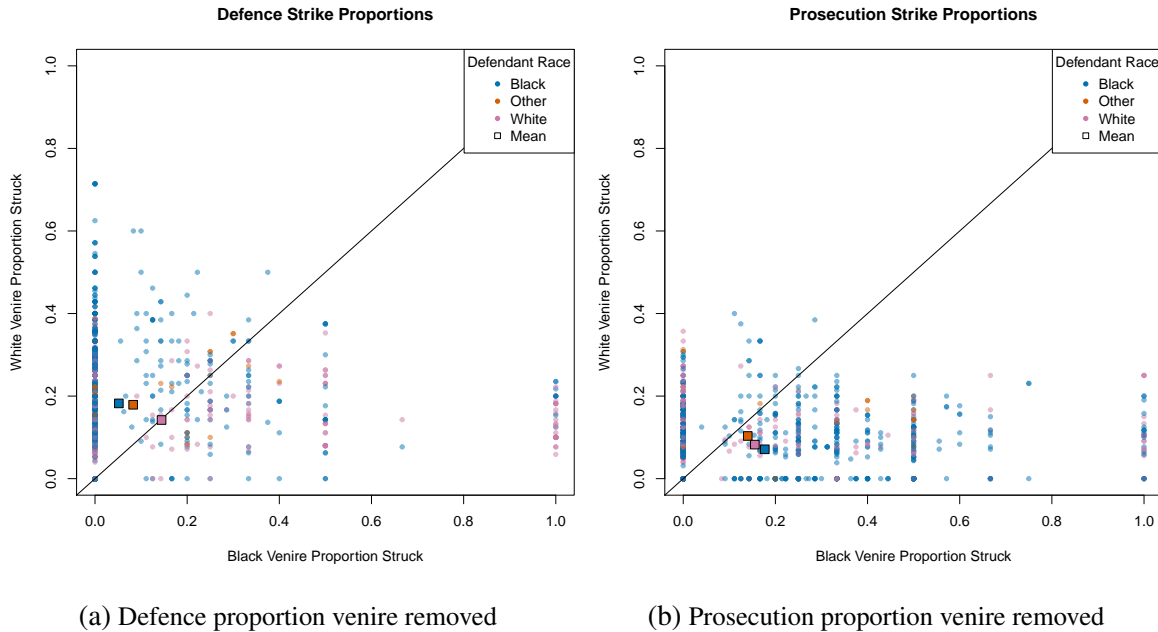(a) Defence proportion venire removed   (b) Prosecution proportion venire removed

Figure 6: Scatterplot of venire proportion removed by race for both the defence and prosecution.

Here, the prosecution and defence biases are 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 observations made in Section 4.2, indicating that these mechanics operate at the trial level, not just the venire member level. More crucially, the high variation visible in Figure 6 suggests that the aggregate patterns described here and in Section 4.2 are not followed by all defence or prosecution lawyers, merely on average.

While these differences are intriguing, a far more interesting observation is the fundamental difference in inclusion of minority and majority groups in jury formation. The aggregate statistics indicate that the black venire members are a minority, and Figure 6 suggests that as a consequence, it is common that a majority or all of the potential black jurors will be removed by peremptory

challenge. There is not a single case where all of the majority white venire members, suggesting the complete removal of white venire members is rare, if not impossible. Whether these peremptories are being used legitimately, the ultimate consequence of their use is the exclusion of minorities from the juries of cases with minority defendants. Given the tarnished history of peremptory challenge use to create identical exclusion in a deliberate manner, this pattern is distasteful to say the least.

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, and an analogous pattern appears to be present in the Philadelphia and Stubborn data sets. 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 Section 4.1, but is visible in the trial summaries presented in Section 4.3. 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 a data set, this work demonstrates the strengths of visual analysis. The first of these is the utility of carefully constructed visualizations 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.

Another strength of the mobile plot is to motivate modelling. The multinomial regression models from Section 4.2 were created to fit models analogous to the displayed patterns of the mobile plots generated to explore the data. That the findings of the models matched the analysis of the mobile plots almost exactly demonstrately the analytical utility of these plots. The models motivated by Figure 1 allowed for the estimation of effects in the Sunshine data controlling for possible legitimate confounders, giving coefficient estimates consistent with those generated previously for other data sets, such as the Stubborn and Philadelphia data.

Visualising these coefficients with the dot-whisker plot made a number of more nuanced patterns obvious. The first of these is the greater sensitivity of the defence to the racial aspects of a trial than the prosecution; 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 decision making across trials.

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

17

strikes 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 minority 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.

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 visualisations 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 Section 4.2 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.

Other possible models of interest are mixed effect models. Such models are attractive because they have the potential to flexibly control for a host of factors which will vary between trials such as judges and lawyers. Estimating a random effect for lawyer, for instance, could shed light on how variable lawyers are in their behaviour. A number of attempts to fit mixed models to this data were made, but all encountered convergence issues. More time spent cleaning and transforming the data may yet bear fruit, however.

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 charges, for example, were not used in the above analysis at all, despite the information content surely contained therein.

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. Such work could complement the visualisations of this paper to facilitate quick and informative comparisons of new data to previous studies.

References

42nd Parliament of Canada (2018). 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).

Aboriginal Justice Implementation Commission (1999). Report of the Aboriginal Justice Inquiry of Manitoba. http://www.ajic.mb.ca/volume.html (Accessed January 2020).

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.

Fine, S. (2019). Jury trials at risk as judge srikes down federal ban on peremptory challenges enacted after Boushie case. Globe and Mail. https://www.theglobeandmail.com/canada/article-ontario-judge-strikes-down-federal-ban-on-peremptory-jury-selection/ (Accessed January 2020).

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).

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. (2018). 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).