

Exploration of Judicial Facial Expression in Videos of Legal Proceedings

A thesis submitted for the degree of
Bachelor of Commerce (Honours)

by

Huize Zhang

27478343



Department of Econometrics and Business Statistics
Monash University
Australia

October 2019

Contents

Acknowledgements	1
Declaration	2
Abstract	3
1 Introduction	1
1.1 Background and motivation	1
1.2 Literature review	1
1.3 Cases selected in the project	3
2 Data Collection	6
2.1 Data processing	6
2.2 Facial variables and action unit	7
2.3 Data format	8
2.4 Missing value imputation	9
2.5 Source code	10
3 Methodology	11
3.1 Notation	11
3.2 Modelling presence	12
3.3 Modelling intensity	14
3.4 Post-model analysis	16
4 Results	18
4.1 Exploratory data analysis	18
4.2 Filtering action units	23
4.3 Modelling result for presence	23
4.4 Modelling result for intensity	26
4.5 Insights	28
5 Conclusion, limitation and future work	30
5.1 Conclusion	30
5.2 Limitation	31
5.3 Future work	31
5.4 Acknowledgement	32

A Appendix	33
A.1 List of videos used in the project	33
A.2 An illustration of face landmarking	34
A.3 Description of action units recognised by OpenFace	35
A.4 Model estimation result	36
A.5 Residual plots	44
A.6 The presence of facial expression of the justices by speaker	45
A.7 The intensity of facial expression of the justices by speakers	45
Bibliography	46

Acknowledgements

I would like to express my gratitude to Professor Di Cook, my supervisors, for detailed guidance and kindness support throughout the project and Professor Russell Smyth for raising the idea of this project. I would like to appreciate Stephanie Kobakian, with whom I have countless discussions with about the project. I would also like to extend my thank to my friends, colleagues and family for standing behind me unconditionally.

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Huize Zhang

Abstract

It is part of human nature to react to change by expressing emotions. However, in some situations it is necessary to attempt to restrict reactions, and expressions of emotions. In many court systems it is required that Justices should restrict emotional displays to ensure the judgement is not biased towards a particular party. In this study, facial recognition software is used to objectively assess the facial expressions of six Justices in six cases (seven videos) heard in the High Court of Australia. Facial information derived from the software is applied in a logistic regression to model the presence score of a selected range of action units. The intensity of the action units is modelled by a two part model. It is observed that Justices generally remain impartial during the court proceedings. Negative emotions such as sadness, fear and anger are associated with action units that occur more intensely or frequently. The requirement to remain expressionless is difficult for some Justices, especially in criminal cases that involve drugs and sexual assaults.

Chapter 1

Introduction

1.1 Background and motivation

The decisions of Justices have always been a source of debate and discussion. Since the realist movement in the United States emerged in the 1930s, many attempts have been made to predict decisions using specific characteristics of the Justices such as gender, political views, and religious backgrounds. More recently, scholars (Shullman, [2004](#); Chen et al., [2016a](#)) have utilised Audio Visual (AV) recordings and transcripts to predict the outcome of cases in the U.S. Supreme Court. Tutton, Mack, and Roach Anleu ([2018](#)) have used an ethnographic approach to study the judicial behaviour. The study involves manually observing the audio footage and taking notes when an obvious emotion is observed. Manually observation may lead to subjective evaluations of facial expressions when different individuals are observing the same AV recordings. This motivates us to extend Tutton, Mack, and Roach Anleu ([2018](#))’s work and employ facial recognition technology to study the facial expressions of the Justices to obtain objective judgements.

1.2 Literature review

The literature summary is divided into two sections: (1) current work in legal studies to understand the behaviour of the Justices and (2) existing facial recognition and emotion tagging technology.

1.2.1 Legal study from a behaviour perspective

There is comprehensive law, economics and political science literature that attempts to predict how the Justices will vote in court cases. Some characteristics of the Justices, for example, gender, political view, religious background has been considered in the literature (Nagel, 1962; Van Koppen and Kate, 1984; Aliotta, 1987; Steffensmeier and Britt, 2001; Kulik and Perry, 2003).

More studies depart from static characteristics of Justices and incorporate the language used by the Justices in the court to understand the decision of the Justices. Black et al. (2011) have studied the use of pleasant and unpleasant language by the Justices. Shullman (2004) and Johnson et al. (2009) have studied the effect of frequency and content of Justices' questions. Epstein, Landes, and Posner (2010) has utilised the number of questions asked by the Justices in regression analysis to predict the winning party in a case .

Recently, legal studies have focused on the usage of emotion and vocal characteristics of the Justices. From a behavioural perspective, Paul Ekman (Ekman et al., 1991) suggests that speakers are often unaware of their own facial and vocal inflections. In 2016, Chen, Halberstam, and Alan (2016) have studied the emotion of the Justices from vocal characteristics and suggest that perceived masculinity is strongly correlated with the court outcomes. Dietrich, Enos, and Sen (2019) also suggest that subconscious vocal inflections contain information that is not available from text transcripts using multilevel logistic model with random effects. Another study by Chen et al. (2016b) have incorporated both vocal and image information into a machine learning model to predict the votes of the Justices, and case outcome, using the U.S. Supreme Court data from 1946-2014. They found that image and audio features have improved the prediction of case outcomes. This demonstrates the potential of incorporating facial information to understand the Justices.

The literature often considers the U.S. Supreme Court Database and far less studies have been conducted using Australian High Court data. The Guide to Judicial Conduct (Chief Justices of Australia and Zealand, 2017) has presented the following code of conduct:

It is important for judges to maintain a standard of behaviour in court that is consistent with the status of judicial office and does not diminish the confidence of litigants in particular, and the public in general, in the ability, the integrity, the impartiality and the independence of the judge.

This highlights the expectation for the Justices to present impartial in the courtroom. Tutton, Mack, and Roach Anleu (2018) has used a novel ethnographic approach to study the judicial demeanour in the High Court of Australia by using transcripts and AV recordings. The study found that Justices present a detached facial demeanour during the court most of the time, but some human display of emotions such as laughter and humour were also captured. However, their approach of manually observing the AV recordings could be biased and lead to subjective results influenced by the individuals observing the videos. An objective approach utilising facial recognition technology is employed in this project to study the expressions of the Justices aiming to produce objective results.

1.2.2 Facial recognition

An anatomical study of the decomposition of facial muscles by Ekman and Friesen (1976) has led to the development of Facial Action Code (FAC), and the identification of the six universal emotions on human faces. This work has laid a solid foundation for analysing facial expressions. Effective facial recognition software, for example DeepFace (Taigman et al., 2014) from Facebook and FaceNet (Schroff, Kalenichenko, and Philbin, 2015) from Google, have been developed for face detection in search and social media platforms. OpenFace (Baltrusaitis et al., 2018) is the first open-sourced face recognition software that provides facial expression detection, including facial landmarking, head pose estimation, eye gaze tracking and facial action unit detection. The OpenFace toolkit has been used in different research areas including depression classification (Yang et al., 2016), emotion studies (Huber et al., 2018) and sports analytics. (Kovalchik and Reid, 2018).

1.3 Cases selected in the project

Six cases have been analysed in this project and they are chosen to cover a broad range of legal areas. Two cases from **immigration** law were chosen because a series of decisions

made by the High Court of Australia related to refugee and immigration status has led the Republic of Nauru to abolish the mechanism that parties could appeal decisions from the Supreme Court of Nauru to the High Court of Australia.

In *Republic of Nauru v WET040* [No. 2] [2018] HCA 60, an Iranian national (respondent) was seeking for asylum protection from the Secretary of the Department of Justice and Border Control (DJBC) and was rejected. Three High Court Justices sat the hearings were Justices Gageler, Nettle and Edelman and the case is referred to as Nauru-a in this project.

Another case from immigration law is *TTY167 v Republic of Nauru* [2018] HCA 61, where a Bangladesh citizen (appellant) applied to Nauru's Secretary of the Department of Justice and Border Control for refugee protection. The appellant then appealed to the Tribunal and further appealed to the High Court of Nauru but was rejected. He then appealed to the High Court of Australia and successfully got his refugee status. This case is also heard by Justices Gageler, Nettle and Edelman and it is referred as Nauru-b.

Rinehart v Hancock Prospecting Pty Ltd [2019] HCA 13 is a **commercial** case discussing commercial arbitration. Due to its complexity, the case were held in two hearings named Rinehart-a and Rinehart-b in the project. Chief Justices Kiefel, Justices Gageler, Nettle, Gordon, and Edelman heard the case. A distinct characteristics of this case is that the decision is not a unanimous decision of all the Justices. Justice Edelman took a narrow interpretation of legal issue while the majority of the Justices took a broad interpretation.

Parkes Shire Council v South West Helicopters Pty Limited [2019] HCA 14 is a **civil** case where the appellant, the Stephenson claimed for psychiatric harm resulting from the death of Mr Stephenson, who was carried and subsequently killed due to the helicopter crash by the Parkes city council (respondent). Chief Justices Kiefel, Justices Bell, Keane, Gordon, and Edelman heard the case and it is referred to as Parkes.

Another two **criminal** law cases are chosen in the project as the nature of criminal cases are highly different from civil cases. In *McKell v The Queen* [2019] HCA 5, which is referred to as case McKell, the appellant is a truck driver and was involved in the importation of drug and cash. The trial judge sentenced a 18 years imprisonment and the appellant appealed to the Court of Criminal Appeal and further to the High Court of Australia. The

High Court Justices Bell, Gageler, Keane, Gordon, Edelman decided there's a miscarriage of justice and quashed the conviction of the appellant.

In *OKS v Western Australia* [2019] HCA 10, the appellant is charged with misconduct with children. The Court of appeal of the Supreme Court of Western Australia charged the appellant for conviction and the appellant then appealed to the High Court of Australia. Justices Bell, Keane, Nettle, Gordon and Edelman unanimously allowed the appeal and issued a new trial.

The links to the hearings of the cases can be found in Table A.1 in the Appendix.

Chapter 2

Data Collection

2.1 Data processing

The Audio Visual (AV) recordings of cases described heard by the High Court of Australia are available on the High Court of Australia website. The workflow to obtain the facial landmarks and expression information from the source videos has been displayed in Figure 2.1.

To download videos from the High Court of Australia the software Youtube-dl (Hsuan, Amine, and Sergey, 2019) is used. Image frames are extracted from each of the videos, at every one minute interval via ffmpeg (<http://www.ffmpeg.org/>), this results in 1021 image frames. The Justices remain seated in the same position throughout the hearings, this means the same region of every image can be extracted to form a set of images containing each individual Justice. Taipan (Kobakian and O'Hara-Wild, 2018) is used to find the x-y coordinates of a box denoting the location of the Justices in each image frame. ImageMagick (Cristy et al., 2019) is used to crop the face of each Justice from each image frame based on the coordinates from Taipan. The resulting 4601 cropped regions containing Justice's faces are then sent to OpenFace (Baltrusaitis et al., 2018) to be processed. The results provided by OpenFace contain facial variables including facial landmarking, head pose, eye gaze and action units. These are stored as separate comma-separated values (csv) files for each of the 4601 faces. Post-processing is done in

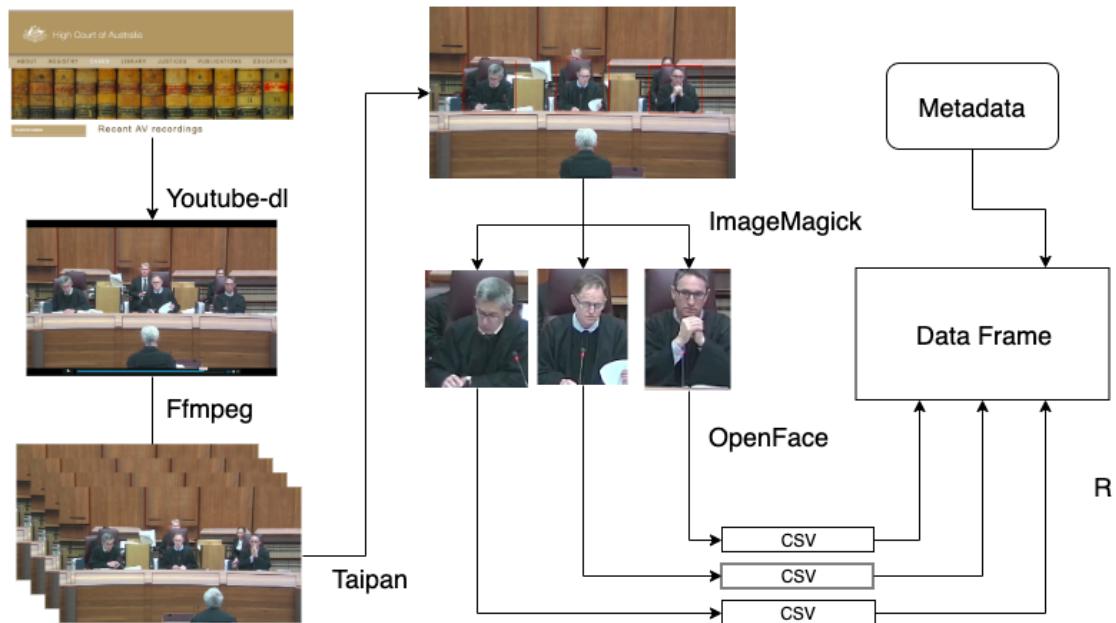


Figure 2.1: An illustration of the workflow for extracting facial variables from videos.

R to combine the separate csv files into a dataframe with additional index columns for frame, judge and video. Lastly, metadata related to the speaking party are extracted from transcript of hearings.

2.2 Facial variables and action unit

OpenFace provides more than 711 variables measuring different aspects of a given face, a full description of the output variables can be found in Baltrusaitis et al. (2018). The facial variables can be summarised into the following categories.

- **Confidence:** How confident OpenFace is in the detection.
- **Gaze:** The vector from the pupil to corneal reflection.
- **Pose:** The location of the head with respect to camera.
- **Landmarking:** The location of certain characteristic points on the face and around the eyes. An illustration of face landmarks can be found in Figure A.1 in the Appendix.
- **Action Unit:** An action unit is used to describe the movement of a single facial muscle.

Human facial expression can be de-constructed into a combination of action units. Happiness is the addition of action unit 6, cheek raiser and action unit 12, lip corner puller. The Facial Action Coding System (FACS) is the common standard for describing facial expressions. To decompose an emotion of sadness, three action units are utilised. Action unit 01 describes the raise of inner brow; action unit 04 is brow lowerer and action unit 15 depicts the lower of lip corner. Action units are chosen to study the facial expressions of the Justices as suggested by Kovalchik and Reid (2018). The action units OpenFace is able to recognise have been provided in Table A.2 in the Appendix.

2.3 Data format

Table 2.1 presents an illustration of the data extracted via the workflow described above in the long format. The presented data is shows the action unit as index adn presence and intensity presented as observations in two columns for Justices Edelman in the first frame of case McKell. Since the frame is cropped at one minute interval, the intensity and presence can also be viewed as time series and Figure 2.2 plots the action unit 1 of Justices Edelman in case McKell across time.

Table 2.1: *An illustration of the data format for Justices Edelman in case McKell for all the action units in the first frame in long format.*

judge	video	frame	speaker	AU	presence	intensity
Edelman	McKell	1	Appellent	AU01	0	0.05
Edelman	McKell	1	Appellent	AU02	0	0.00
Edelman	McKell	1	Appellent	AU04	0	0.01
Edelman	McKell	1	Appellent	AU05	0	0.00
Edelman	McKell	1	Appellent	AU06	0	0.00
Edelman	McKell	1	Appellent	AU07	0	0.00
Edelman	McKell	1	Appellent	AU09	0	0.26
Edelman	McKell	1	Appellent	AU10	0	0.00
Edelman	McKell	1	Appellent	AU12	0	0.00
Edelman	McKell	1	Appellent	AU14	1	1.23
Edelman	McKell	1	Appellent	AU15	0	0.46
Edelman	McKell	1	Appellent	AU17	0	0.66
Edelman	McKell	1	Appellent	AU20	1	1.44
Edelman	McKell	1	Appellent	AU23	0	0.64
Edelman	McKell	1	Appellent	AU25	0	0.00
Edelman	McKell	1	Appellent	AU26	0	0.00
Edelman	McKell	1	Appellent	AU45	0	0.25

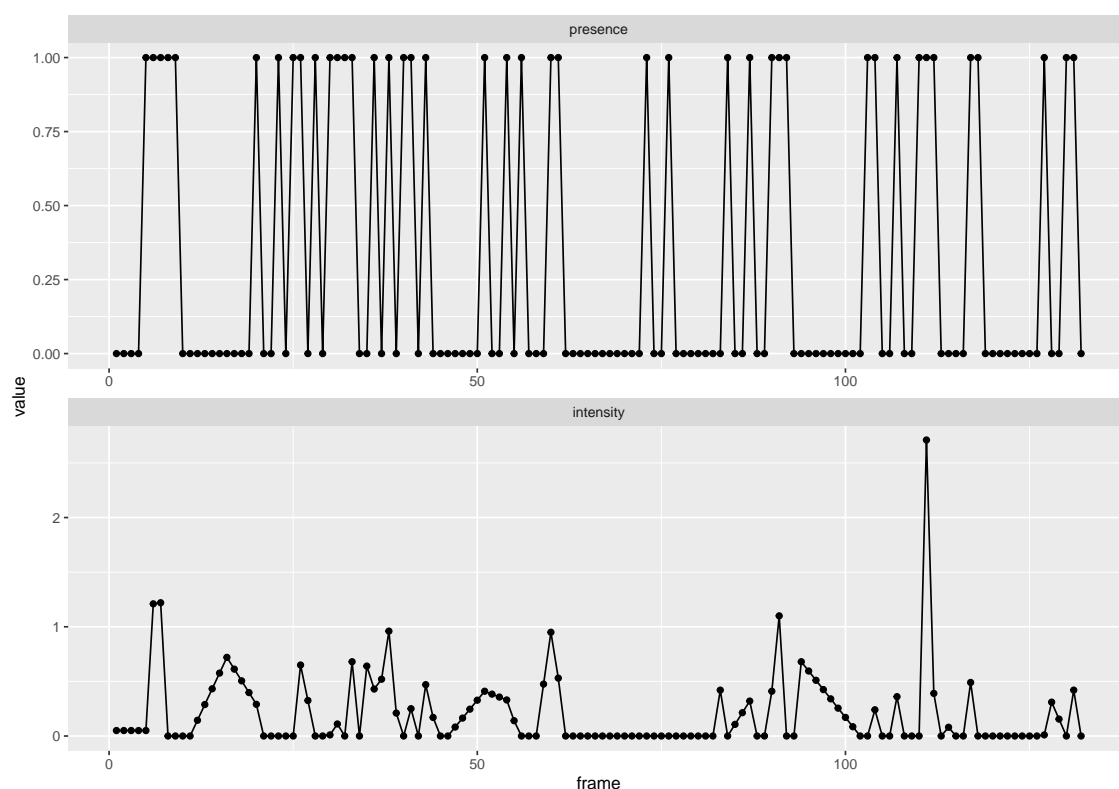


Figure 2.2: *The intensity and presence score of action unit 01 for Justices Edelman in case McKell is graphed against time (frame number) as line chart. The intensity is a numerical variable while presence is binary variable takes value of 0 when the action unit is not present and 1 otherwise.*

2.4 Missing value imputation

Missing values occur in the data whenever the Justice is not looking straight ahead. This might occur when they are reading materials on their desk, or perhaps if conversing with their legal assistant behind them. It can also occur when there are five Justices on a case, and the video resolution is not sufficiently high to detect the face or action units. The data structure that we created specifically places an NA in these positions. This has allowed us to examine the pattern of missings and check it happens more often in some recognisable way, for example when an appellant is speaking. We did not find any over-arching pattern, and thus have used a simple procedure to impute missings for intensity, which was then used to impute presence.

Intensity is a continuous variable ranging from zero or five measuring how strong the action unit is presented. Linear interpolation function (`na.interp()`) from forecast

package is used to impute Intensity. The missing value of presence is then imputed as one if the intensity score of the missing observations are greater than one and zero otherwise.

2.5 Source code

Source code for the workflow of data processing is available at https://github.com/huizezhang-sherry/ETC4860/data_pre_processing. The full data obtained from OpenFace after post-process is named `full_data` available in the `raw_data` folder in the same repository. The imputed dataset contains only action unit is named `au_imputed` and also available in the `raw_data` folder.

Chapter 3

Methodology

3.1 Notation

Let \mathbf{X} be a matrix of predictors, and \mathbf{Y} variable is a bivariate matrix of response variables, including a binary indicator of presence/absence and a numeric value measuring intensity, of facial action unit, where

- X_1 indicates judge with six categories $i = 1, 2, \dots, 6$
- X_2 indicates video for each of the seven cases, $j = 1, 2, \dots, 7$
- X_3 indicates action unit containing 18 possible facial expression.
- X_4 indicates speaker, either the appellant or respondent, $l = 1, 2$
- X_5 indicates frame corresponding to time, $t = 1, 2, \dots, T_j$

Note that t could be considered a time variable, but because images are taken at 1 minute intervals, temporal dependence is unlikely to exist. Rather this should be considered an independent observation.

A full, main effects model for the data might be expressed as:

$$Y_{ijklt} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + \varepsilon_{ijklt}$$

Also, let P_{ijklt} represent the response variable presence, and I_{ijklt} represent the response variable intensity. This notation will be helpful for defining the plots and models explained in this section.

3.2 Modelling presence

3.2.1 Model structure

The presence score is a binary variable that is one when a particular action unit is observed and zero if not. A logistic model is able to capture this binary feature of the presence score and we implement it using the `glm()` function from base R. The link function of a matter of choice in the generalised linear model and the logit link is chosen because it is the canonical link of the binomial family. An alternative link could be a probit link but theoretically, these two links give very similar result in terms of prediction (Faraway, 2016). The structure of the model is written in Equation 3.2 with the first equation linking the mean of the presence to the linear prediction and the second equation specifying the linkage between η to predictors. The next section will specify three different function forms for the linear predictor.

$$\mu = \frac{e^\eta}{1 + e^\eta} \quad (3.1)$$

$$\eta = f(\alpha_i, \beta_j, \gamma_k, \delta_l) \quad (3.2)$$

3.2.2 Model 1: Action unit

The first linear predictor is presented in Equation 3.3. It includes the main effect of judge, action unit and also their interaction. Interaction terms are included to capture the judge-wise differences for different action units and it is necessary because we suspect different judges could have different average presence scores for different action units.

$$\eta_{ik} = \mu + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} + \varepsilon_{ik} \quad (3.3)$$

3.2.3 Model 2: Video

Build upon the first model, the second model adds the video related main effect and interactions, as shown in Equation 3.4. The interactions allow both variable judge and action unit to differ in different videos. This model structure is useful to answer the research questions *Whether the judges are behaving same or different across videos?*

$$\eta_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijk} \quad (3.4)$$

3.2.4 Model 3: Speaker

Build upon the second model, the third model is aimed to capture the speaker-wise effect by including the judge and speaker interaction as in Equation 3.5. This model is built attempting to answer the question: *Do the expressions of the judges change when different parties are speaking?*

$$\eta_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\delta)_{il} + \varepsilon_{ijkl} \quad (3.5)$$

3.2.5 Model comparison

The analysis of variance (ANOVA) (Faraway, 2016; Gelman and Hill, 2006) is a statistical method for model comparison. We use the base R `anova()` function to compare the three models via chi-square tests.

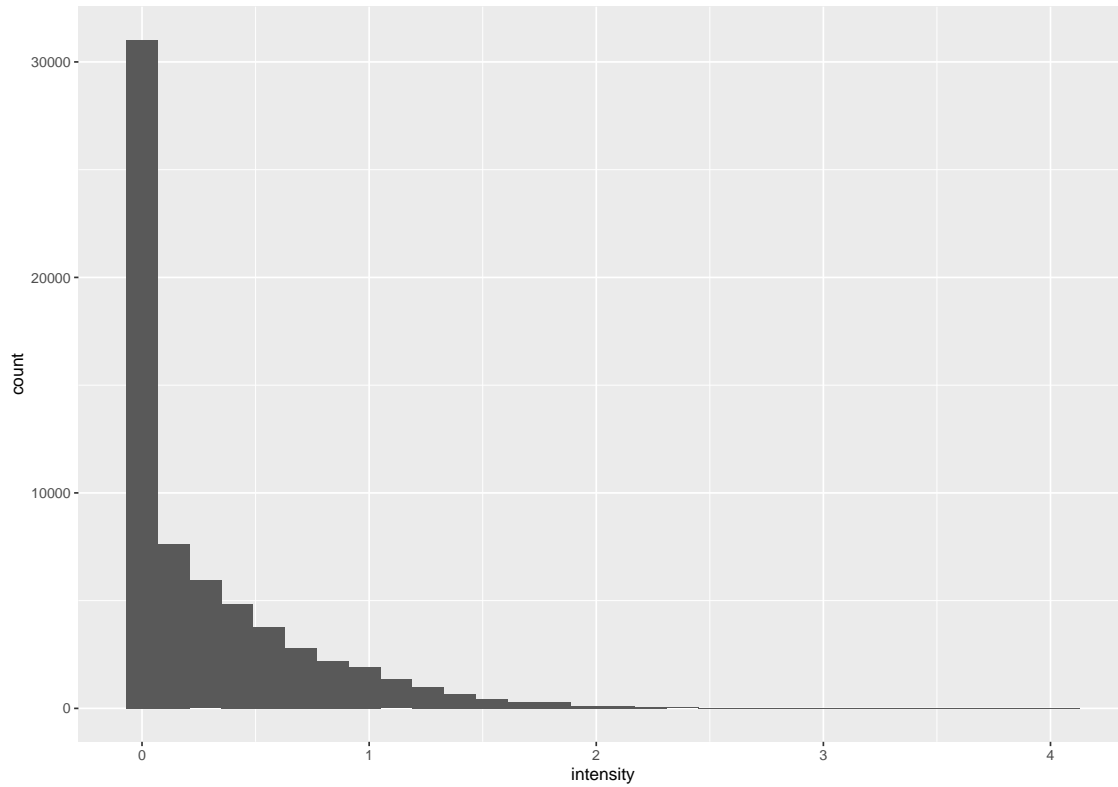


Figure 3.1: *From the histogram of the intensity score, the data is highly skewed with an excessive amount of zeros. The two part model is about to accommodate the excessive zeros via the logistic model and gamma regression is about to capture the skewness in the data.*

3.3 Modelling intensity

The intensity score is a continuous variable, with zero indicating an action unit is not present to a maximum intensity of five and a histogram of the intensity is plotted in Figure 4.3. The data has a high proportion of zeros and the non-zero values are highly skewed. This type of data is the so-called semi-continuous data (Liu et al., 2010). and can be modelled in the econometrics literature by a two part model (Cragg, 1971). In the two part model, the data is viewed to be generated sequentially, which has a mixed distribution of

- a logistic model of if $Y = 0$ or not, and
- a specific model for the conditional distribution of $y \mid y > 0$.

The choice of model between two part model and sample selection model is always discussed in the literature. Monte-Carlo simulation studies by different researchers (Leung

and Yu, 1996; Duan et al., 1984; Manning, Duan, and Rogers, 1987) show different results on whether these two classes of model are answering the same or distinct inferential questions. The reason for us to choose two part model rather than sample selection model is because unobservability is not a problem in our case. In another word, if an action unit is not present for an observation, it doesn't make sense to talk about "intensity score if the action unit was present". Tobit model is not appropriate because the data can't be viewed as normally distributed with negative value censored as zero (meaningless to say negative intensity value). Zero inflated model is not used because it considers two source of zeros in the data while there is no zeros being generated from the conditional distribution defined below.

The two part model has a general structure as in Equation 3.6.

$$\mu^1 = \frac{e^\eta}{1 + e^\eta} \quad (3.6)$$

$$\eta = f(\alpha_i, \beta_j, \gamma_k, \delta_l) \quad (3.7)$$

$$\mu^2 = \log(I) \quad (3.8)$$

$$E(I \mid I > 0) = f(\alpha_i, \beta_j, \gamma_k, \delta_l) \quad (3.9)$$

where μ^1 is the mean of the intensity score and μ^2 is the mean for intensity given intensity > 0 . The first two equations capture the logit link and its linear predictor in the logistic regression. The third and fourth equation specify the functional form of the conditional distribution.

The functional form of the conditional distribution need to be able to capture the highly skewed nature of the non-zero observations. A convention approach is to assume the conditional distribution is a lognormal distribution (Diehr et al., 1999). More recent literature proposes the use of gamma or generalised gamma regression model (Liu et al., 2010). Gamma regression is chosen to because it could also capture the right skewness and it can be easily implemented via the `glm()` function. The log link is used in the gamma

regression because the canonical inverse link will cause some estimated marginal means to be extremely high and thus meaningless for intensity score.

The linear predictor of the conditional intensity that includes video and relevant interactions is written in Equation 3.10.

$$E(I_{ijk} \mid I_{ijk} > 0) = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \quad (3.10)$$

The model that captures additional speaker variable is written in Equation 3.11.

$$E(I_{ijkl} \mid I_{ijkl} > 0) = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\delta)_{il} \quad (3.11)$$

3.4 Post-model analysis

The estimates of variables from the model summary are not particularly useful for our purpose. This is because firstly, the estimates of the coefficients are not interpretable in the logistic regression. Secondly, we are interested in *whether the mean for each treatment is same or different?* To assess which level of the factor is different requires post-model analysis.

3.4.1 Estimated Marginal Mean (EMM)

The estimated marginal mean (Gelman and Hill, 2006) is the fitted value from a model over the treatment effects. The treatment effects include judge, video and action unit in Model 2 and an additional speaker in Model 3. The estimated marginal mean is computed using `emmean()` from the `emmenas` package. The probability from estimated marginal mean can be interpreted as the estimated probability of presence (and intensity) score for a particular combination of action unit, judge and video. This output allows us to compare how the estimated presence (and intensity) are different or similar from each other.

3.4.2 Confidence Interval Adjustment

The confidence intervals computed from the `emmean()` function need to be adjusted for simultaneous inference. A 5% significance level indicates if we conduct 100 tests simultaneously, about 5 tests will show significance out of randomness. This is a problem we need to pay attention to when comparing the estimated presence probability or we may wrongly conclude judges has a different facial expression than others but they are actually not.

When multiple estimated mean are compared at the same time, the confidence level need to be adjusted to control the family-wise error rate to be less than α . Bonferroni adjustment makes the adjustment to reject a hypothesis test at α / N to control for the Family-wise Error Rate. `Confint()` function from base R is used with argument `adjust = "bonferroni"`.

Chapter 4

Results

4.1 Exploratory data analysis

4.1.1 Action unit: presence

Mean presence score and most common action units

Follow the notation defined in Chapter 3, the average presence score (P_{ik}) of each action unit is computed for each judge as

$$P_{ik} = \frac{\sum_{jt} X_{ijtk}}{\sum_{j=1}^J T_j}$$

Figure 4.1 graphs the presence score of all the action units across all the judges. The order of action unit on the y axis is ranked by the average presence of all the judges. The five most frequent action units are highlighted in blue. From Figure 4.1, some of the action units are common across almost most of the Justices, these includes AU02 (outer eyebrow raiser), AU20 (lip stretcher), AU15 (lip corner depressor), AU01 (inner brow raiser) and AU14 (dimpler). Relating to emotions, AU01 and AU15 contribute to sadness. AU02, outer eyebrow raising, can be associated with surprise, fear or interested. Dimpler (AU14) could be linked to contempt or boredom and Action unit 20, Lip Stretcher, is commonly

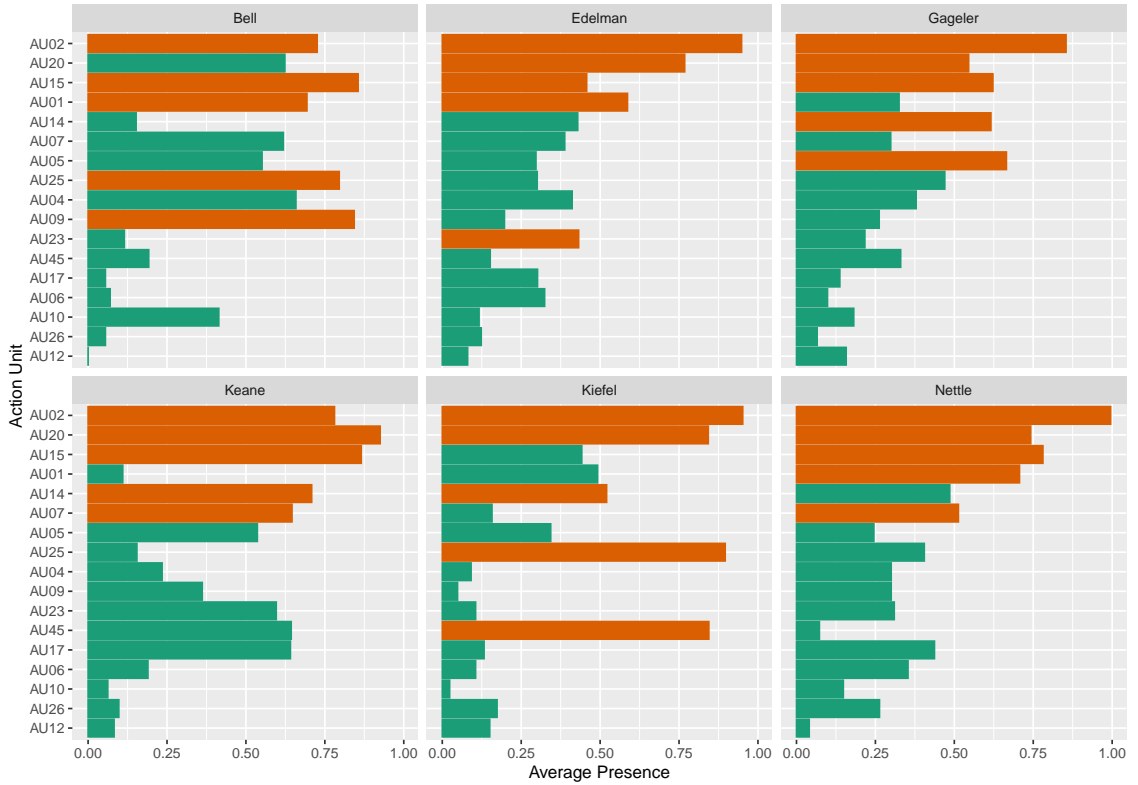


Figure 4.1: The average presence score of each action unit for each Justice, aggregating on video and time. The most common five action units for each Justices is colored in blue. The most common action units across all the Justices include AU02 (outer eyebrow raise), AU20 (lip stretcher), AU15 (Lip Corner Depressor), AU01 (Inner brow raise) and AU14 (Dimpler)

contribute to fear, which is most sophisticated emotion that requires seven separate action units to describe.

Presence by videos

The main presence score of the judges by video (P_{ijk}) is computed as

$$P_{ijk} = \frac{\sum_t X_{ijtk}}{T_j}$$

for the four most common action units: AU02, AU14, AU15, AU20 and presented in Figure 4.2. From this figure, AU02, outer eyebrow raise, appears consistently highly across Justices and court cases. The other three vary across both Justices and cases. AU15, lip corner depressor, varies across Justices: it is common in Justices Bell, Keane and Nettle, but less common in Justices Keane and Edelman. Justice Gageler varies a lot in usage across

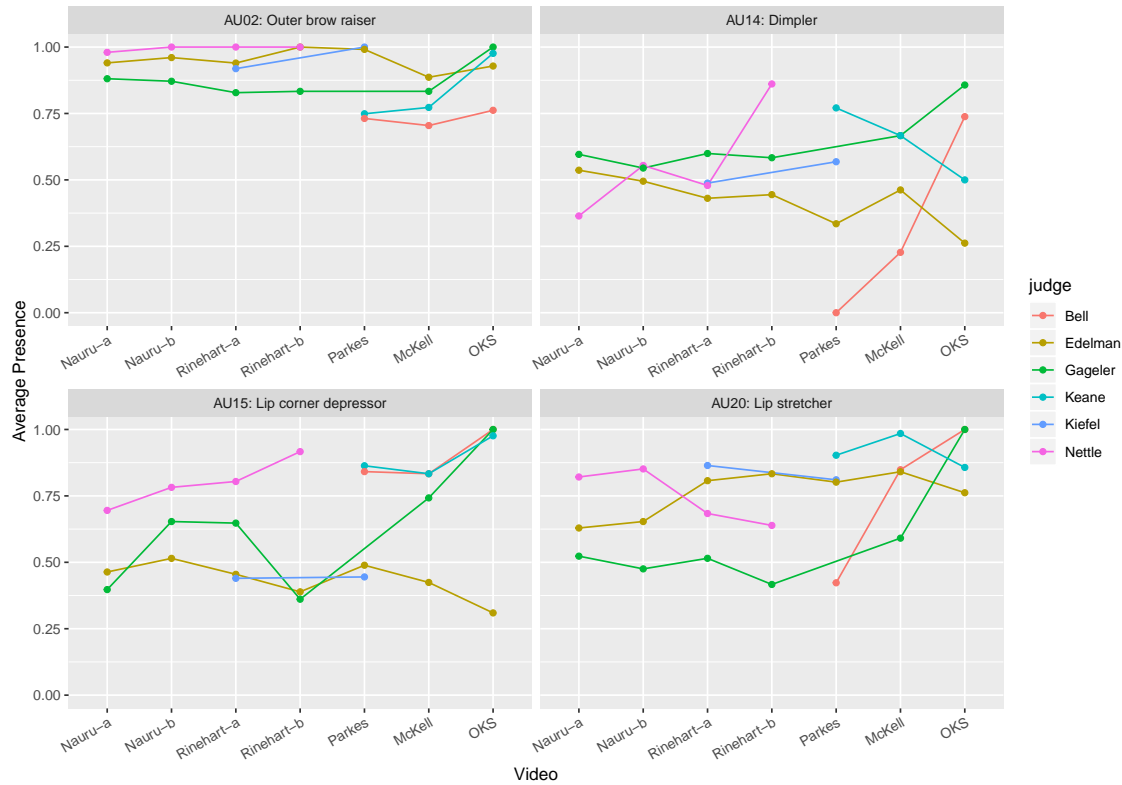


Figure 4.2: Average presence of the four most common action units for each judge by videos. Some Justices, for example Justices Gageler and Bell show large fluctuation on their facial expressions while others are not.

cases and particularly uses this expression in OKS. AU20, lip stretcher is consistent across cases, varies by Justices, but is particularly frequently used in the OKS case by Justices Bell and Gageler. AU14, dimpler, is similar to AU20. Most reactions appear to be happening in case OKS and McKell. Recall that OKS is a criminal case involving misconduct with children, the result above provides some exploratory evidence that the Justices react more frequently in criminal cases like OKS and McKell.

4.1.2 Action unit: intensity

General intensity plot

The boxplot of the intensity for all the Justices across all the videos is presented in Figure 4.3. Each bar-and-whisker represents the intensity (I_{ijt}) of all the action units aggregated on time for a particular Justices i in a specific case j . For example, the first bar-and-whisker in case Nauru_a is created using all the action units of Edelman throughout the elapsed

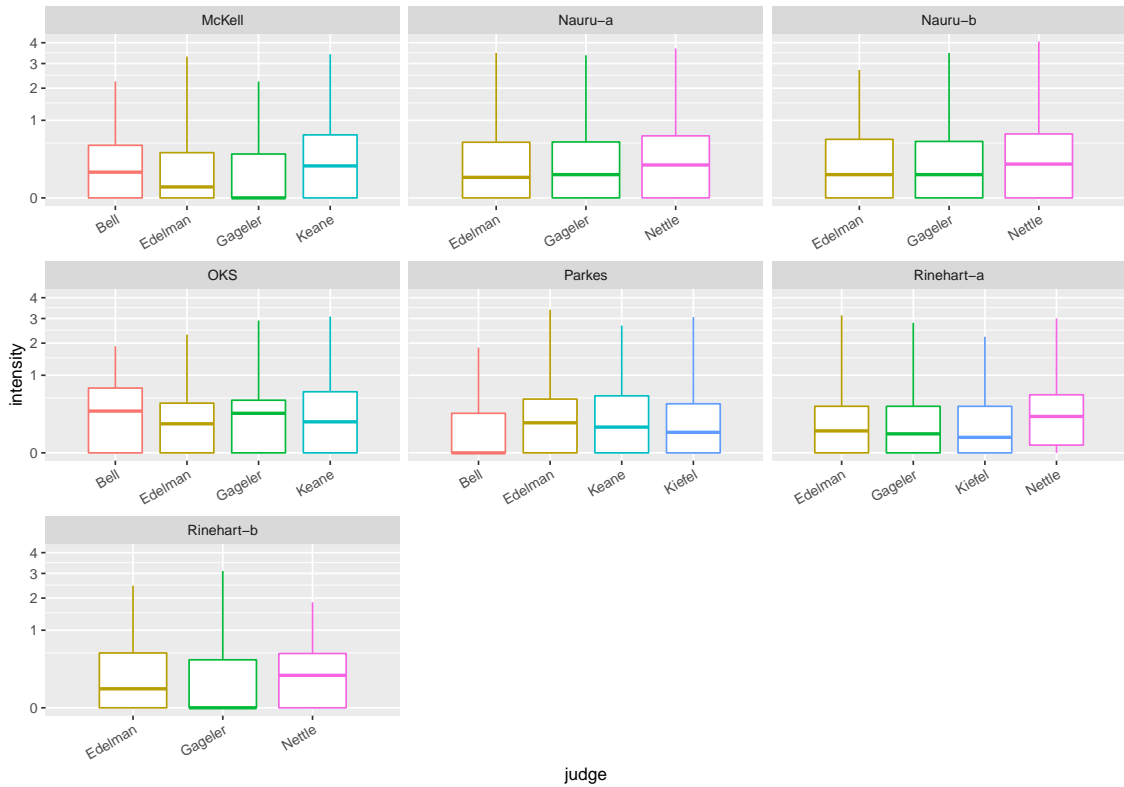


Figure 4.3: Boxplot of intensity score by Justices and videos. Square root transformation is taken since the mean intensity scores are all below one.

time in Nauru_a case. The square root transformation is applied to make the mean easier to be visualised. Most of the action units have low intensity score as shown in the figure, which matches with the prior belief that the Justices are expected not to express too much of their expressions in the courtroom. Justice Nettle, colored in pink has the highest average in all the four cases he appeared.

High intensity points

The points with intensity greater than two are shown against time for all the justices in Figure 4.4. Justices Edelman, Gageler and Nettle are the judges who have stronger expressions that can be detected since they have more points with intensity greater than two. Different Justices also have different time where they display stronger emotions. For example, Justice Edelman is more likely to have stronger emotion throughout the time while Justice Nettle is more likely to have intense facial expressions at the beginning and ending of the hearing.

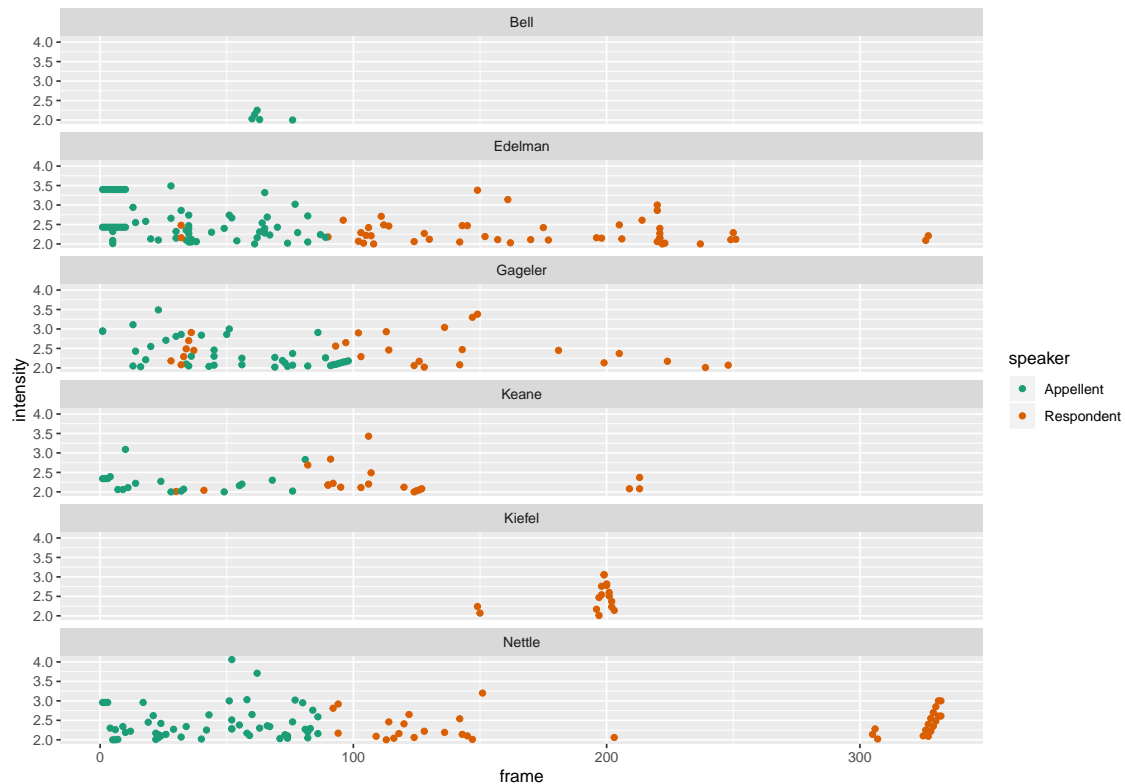


Figure 4.4: Points with intensity greater than two are plotted against time, colored by speaking parties. Justices Edelman, Gageler and Nettle have more intense expressions than other Justices. Justice Nettle has a clear cut on when he is likely to express stronger expressions.

4.1.3 Summary

The findings from the exploratory data analysis are summarised below:

- The most commonly presented action unit from the Justices are AU02 (outer eyebrow raiser), AU20 (lip stretcher), AU15 (lip corner depressor) and AU14 (dimpler).
- Some Justices show relatively consistent facial expression through different videos while others, for example Justices Gageler and Bell have larger fluctuation on their facial expressions in different cases.
- The overall intensity of the action units are low while Justices Nettle has a relatively higher mean intensity than other Justices.

- Edelman, Gageler and Nettle are the Justices with more intense facial expressions in the courtroom and Justice Nettle is the only Justice that tends to have stronger expression towards the end of the hearing.

4.2 Filtering action units

The number of action unit to include in the model is a matter of choice. The discussion of this choice is to ensure the model is parsimonious, that is, a model has the smallest number of variables but with greatest explanatory power. Random effect is a way to deal with large number of factor levels of a variable, but in our context, we are only interested in the action units with a certain mean presence and intensity for most of the judges.

The mean presence and intensity score for each action unit is computed and the action units to include in the model are the ones that appear in the top 10 action unit in both mean presence and intensity rank. This ensures that these action units have both relatively high intensity and presence score. A list of included action units along with their meaning and related emotions are presented in Table 4.1

Table 4.1: *These are the selected action units that will be included in the modelling for intensity and presence.*

AU-number	meaning	emotion
AU01	Inner brow raiser	sadness, surprise and fear
AU04	Brow lowerer	sadness, fear, anger and confusion
AU05	Upper lid raiser	surprise, fear, anger and interested
AU07	Lid tightener	fear, anger and confusion
AU14	Dimpler	contempt or boredom if appears unilaterally
AU15	Lip corner depressor	sadness, disgust and confusion
AU20	Lip stretcher	fear
AU45	Blink	NA

4.3 Modelling result for presence

4.3.1 Model comparison

The three models in Equation 3.3, 3.4 and 3.5 have been fitted and ANOVA test is performed to choose the best model. The ANOVA result for comparing Model 1 and 2 is

presented in Table 4.2. After incorporating the main effect of case and its interaction with judge in Model 2, the degree of freedom is reduced by 61. This has a significant improvement on the model since the p-value ($1e-88$) is close to zero, indicating the null hypothesis that Model 1 and Model 2 are the same is rejected.

The ANOVA result between Model 2 and Model 3 is presented in Table 4.3. The additional six variables associated with speakers in Model 3 have improved the model at 95% significance level since the p-value is 0.05, however, at 99% significance level, this improvement is not significant. Model 2 is chosen as the final model because the interpretation of video-wise effect using Model 2 after post-model analysis provides more interesting findings about the expressions of the Justices than the speaker-wise effect using Model 3.

Table 4.2: *Model comparison using ANOVA for Model 1 and 2. The inclusion of video related variables in Model 2 decreases the degree of freedom by 61 while provides a significant improvement on the model as indicated by the p-value.*

Model	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Model 1	30310	30322.65			
Model 2	30249	28407.13	61	1916	0

Table 4.3: *The model comparison result of Model 2 and Model 3 using ANOVA. The inclusion of six speaker related main and interaction effects contribute to improve the model. However, this is not significant at 99% significant level. Model 2 is chosen as the final model because the interpretation of video effect provides more interesting findings about the facial expressions of the Justices.*

Model	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
Model 2	30249	28407.13			
Model 3	30243	28394.54	6	13	0.05

4.3.2 Residual Diagnostics and post-model analysis

The residuals of Model 2 are plotted against variable judge and video in Figure A.2 and A.3 in the Appendix. There is no obvious pattern shown in the residuals for different Justices or videos, which indicates adequate fit.

The estimated marginal mean is computed and presented in Table A.3 in the Appendix due to its length. The prob column can be interpreted as after averaging over all the

videos and speaking parties, the estimated mean probability for judge Edelman in action unit AU02 is 0.95, with a 95% confidence interval of [0.92, 0.97]. Notice that confidence intervals for a generalised linear model is asymmetric around the estimates because the linear symmetric interval of the mean has been transferred via the inverse of link function to get the confidence interval for the response.

4.3.3 The presence of facial expression of the justices by video

The 95% confidence interval after bonferroni adjustment is plotted in Figure 4.5. In general, most of the intervals for the same judge in the same action unit are overlapping with each other on the vertical axis, while there are some non-overlappings highlight the potential inconsistency of the facial expressions of the Justices.

Justice Edelman and Keane behave consistently throughout all the videos, while they both seem to express significantly less in action unit 5 (upper lid raiser) in the OKS case. Justice Nettle has relatively low expression of action unit 4 (brow lowerer) in case Rinehart-a. Gageler shows a consistently high number of expressions in case OKS for action unit 15 (lip corner depressor) and action unit 20 (lip stretcher).

Bell presents similar reactions to Gageler, showing a significantly higher proportion of emotions associated with action unit 1 (inner brow raiser), 14 (dimpler), 15 (lip corner depressor) and 20 (lip stretcher) in case OKS. Bell also exhibits less presence of action unit 07 (lid tightener) and 20 (lip stretcher) in case Parkes.

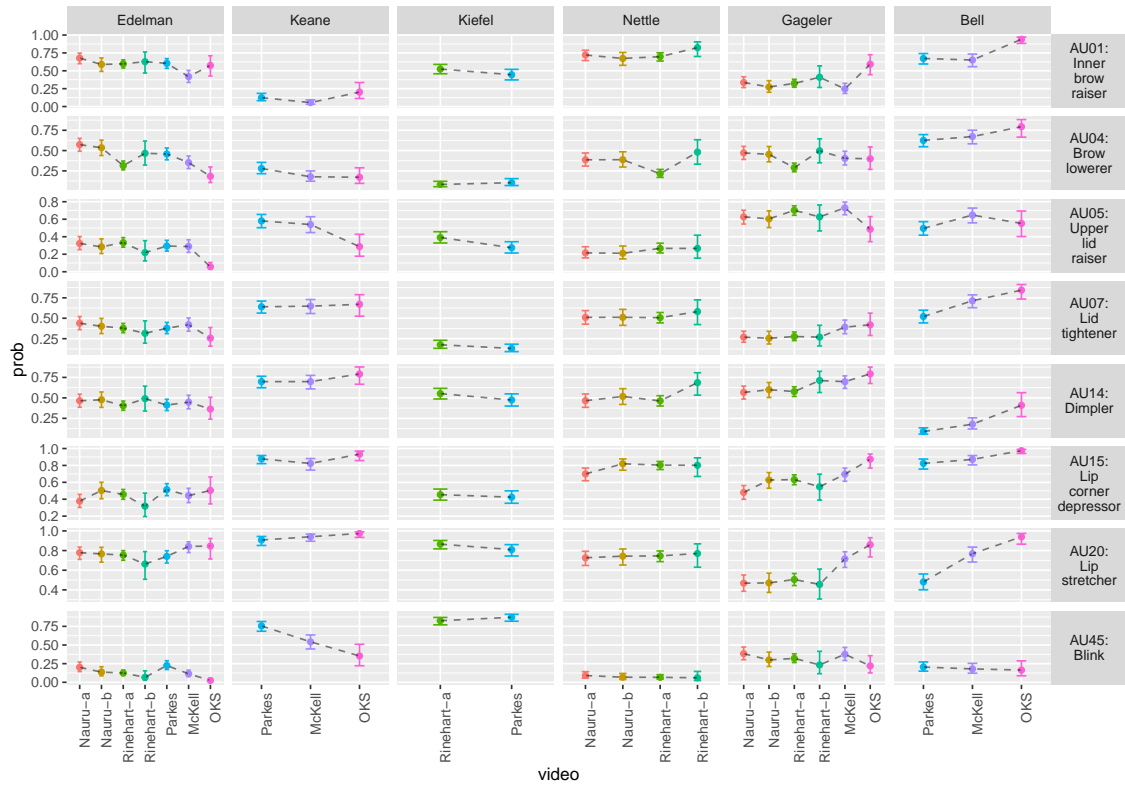


Figure 4.5: The 95% confidence interval for estimated mearginal mean for presence after bonferroni adjustment. The x axis represents video and the y axis represents the estimated marginal mean of an action unit being observed. The facet shows the Justices in columns and action units in rows.

4.4 Modelling result for intensity

4.4.1 The intensity of facial expression of the justices by video

The two part model in equation 3.10 is estimated for the intensity data. Estimated marginal mean and confidence interval adjustment procedure are performed as modelling presence data. The 95% confidence interval plot is presented in Figure 4.6. This shows that Justices Edelman has significantly stronger expressions of brow lowerer (AU04) in case Nauru-a, Nauru-b and Rinehart-a, but less intensity when expressing lid tightener (AU07) in case OKS. Justice Keane also shows more intense expressions of lid tightener (AU07) in case McKell.

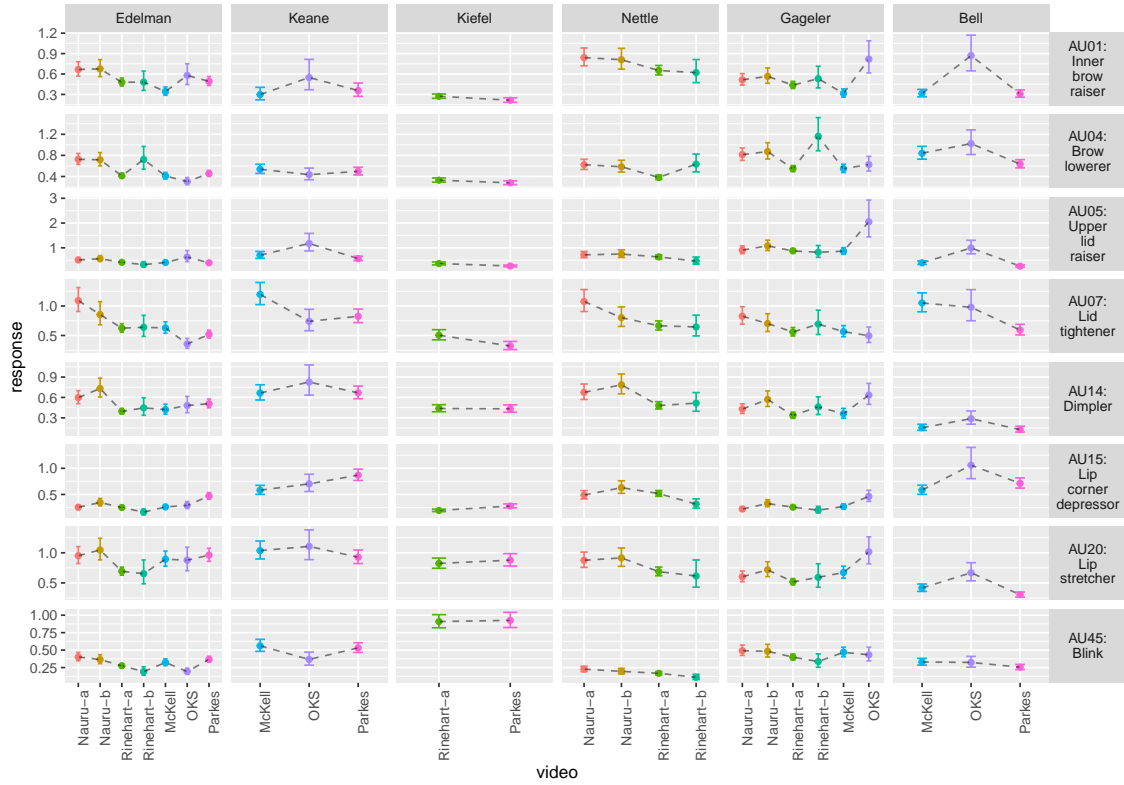


Figure 4.6: The 95% confidence interval for estimated mearginal mean for presence after bonferroni adjustment. The x axis represents video and the y axis represents the estimated marginal mean of the intensity. The facet shows the Justices in columns and action units in rows.

Action unit 5 (upperlid raiser) and 20 (lip stretcher) are exhibited significantly more intense for Justices Gageler in case OKS. The mean for brow lowerer (AU04) seems to higher than those in other cases for Justices Gageler but this result is not significant.

For Justice Bell, the intensity of inner brow raiser (AU01), upper lid raiser (AU05), dimpler (AU14) and Lip stretcher (AU20) are also significantly higher in case OKS.

4.4.2 The expression of the justices by speaker

From the presence and intensity figures which are colored by speakers in Figure ?? and A.4 in the Appendix, we can observe that the video-wise difference between Justices is still preserved when the speaker effects are included in the model. However, the speaker-wise difference is not significant in terms of both presence and intensity for all the Justices.

4.5 Insights

This result from previous chapter contributes to answer the question: **For the same judge, does the mean presence and mean intensity of the action units stay the same or vary for different videos?** In general, the facial expressions of the justices appear impartial, as most of the 95% confidence intervals for the same judge and action unit overlap in the vertical direction in most of the videos in both figures. There are some instances when in a particular video, a judge expressed significantly more or less of an expression.

It is necessary to link back to the nature of the cases to interpret the facial behaviour of the Justices. Nauru-a and Nauru-b discusses immigration law; McKell and OKS are more criminal cases; Parkes is a civil negligence case and Rinehart-a and Rinehart-b are commercial cases arguing contract arbitration.

Based on the nature of the cases, Justice Edelman is more likely to express stronger emotion to the immigration and commercial cases and express less and softer at criminal cases. The action unit 5 (upper lid raiser) and 7 (lid tightener), which are expressed more frequent and more intense by Justice Keane in case OKS are usually associated with the emotion of anger. This implies that Justices Keane is more responsive to the criminal cases. Kiefel and Nettle are relatively consistent in their expressions. Of the six universal emotions, the action units Justices Gageler have significantly more frequent and stronger in case OKS are action unit 5 (upper lid raiser), 15 (lip corner depressor) and 20 (lip stretcher). These three action units are commonly associated with anger, sadness and fear respectively, which indicate Justice Gageler's strong and frequent emotional responses when hearing criminal cases OKS. The result for Bell suggest the same emotional reaction as judge Gageler to criminal cases.

As the speaker-wise difference was not significant, suggestions that Justice favour an appellant or respondent were not confirmed. This result would be a validation that on the high court level, the judges are behaving impartial to different speaking parties.

To summarise, the above discussion of intensity and presence of action unit in different cases gives us several findings about the expression of the judges:

- 1) In general, the expression of the Justices are impartial, which is live up to the code of conduct from Chief Justices of Australia and Zealand (2017) and validate the result from Tutton, Mack, and Roach Anleu (2018).
- 2) When there is significantly present or intense expression of the Justices, it tends to be associated with negative emotion like sad, fear and anger. This could have implication on the mental well-being of the judges.
- 3) Some justices, for example Keane, Gageler and Bell are more responsive, both in frequency (mean presence) and magnitude (mean intensity) to criminal cases. This could show that it is harder for judges to keep a still face when the content of a case goes against human nature.

Chapter 5

Conclusion, limitation and future work

5.1 Conclusion

In this thesis, we explore the facial expressions of seven high court Justices in six cases utilising the publicly available videos from the high court hearing. The main aim of this research is to use a statistical and objective approach to understand whether the Justices are behaving impartial in the courtroom.

Our approach involves extracting facial variables from the videos of the high court hearings and statistically model the presence and intensity of the action units. This allows us to understand whether different Justices would have variations in their expressions in different cases and whether their expressions will be different when different parties are speaking. We have found that in general, the Justices are behaving impartial during the court, which is a validation on Tutton's ethnographic study on the same topic. We also find that Justices tend to have stronger and more frequent negative emotions, for example sad, anger and fear in criminal cases. From a humanity perspective, it could be hard for the Justices control their expressions in criminal cases when extreme and violent scenes are described in the hearing.

One of highlights of the project is to establish a workflow for systematically extracting facial variables from videos. The established workflow makes it easy for any re-processing of the videos and analysing facial expressions from other video source. Furthermore, as far as we know, this study is the first of its kind to statistically analyse videos to study the emotions in the courtroom. This piece of work therefore makes a significant contribution to the legal research by providing a new, statistical methodology to understand the emotion of the Justices. The facial information gained from this research could also be incorporated with other judicial information to predict the high court case outcome in Australia.

5.2 Limitation

The current image frames are extracted at every one minute interval. However, some facial expressions may only last for a few second. Thus more frequent time interval could be used for getting more precise facial information of the judges. Also, if videos of the high court hearing could be accepted as input for facial expression detection, the potential correlation of emotion could be captured even better.

In my work, seven videos are being processed into the facial recognition software and more videos could be processed to get more robust results. The reason for not processing more videos in the current study is because the resolution of publicly available videos from the high court has only 720 pixels while the facial recognition software, OpenFace requires at least 30 pixels for a face to be detected. This means that we have to choose videos where three or five judges are presented.

However, this work has established a workflow for extracting facial expressions of human from videos. As long as more higher resolution videos are available, facial variables can be extracted via the same fashion.

5.3 Future work

Faces could be extracted more often than at 1 minute intervals to allow researchers to capture more precise expressions of the judges. However, as the extraction becomes more

frequent, the problem of serial correlation could rise and appropriate modelling technique should be utilised to accommodate for this feature of data.

5.4 Acknowledgement

The analysis is conducted using R (R Core Team, 2019), and the following packages: forecast (Hyndman and Khandakar, 2008), tidyverse (Wickham, 2017), emmenas (Lenth, 2019) and broom (Robinson and Hayes, 2019). This thesis document is created with knitr (Xie, 2014), R Markdown (Xie, Allaire, and Golemund, 2018) and bookdown (Xie, 2016). All materials required to reproduce the project can be found at <https://github.com/huizezhang-sherry/ETC4860/>.

Appendix A

Appendix

A.1 List of videos used in the project

Table A.1: *Details of videos processed.*

Case	Name	AV recording link	Judge
The Republic of Nauru v WET040 [No. 2] [2018] HCA 60	Nauru_a	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-07a	Nettle, Gageler, Edelman
TTY167 v Republic of Nauru [2018] HCA 61	Nauru_b	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-07b	Nettle, Gageler, Edelman
Rinehart v Hancock Prospecting Pty Ltd [2019] HCA 13	Rinehart_a	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-13	Gordon, Gageler, Bell, Keane, Edelman
Rinehart v Hancock Prospecting Pty Ltd [2019] HCA 13	Rinehart_b	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-14a	Gordon, Keane, Bell, Gageler, Edelman

Case	Name	AV recording link	Judge
Parkes Shire Council v South West Helicopters Pty Limited [2019] HCA 14	Parkes	http://www.hcourt.gov.au/cases/cases-av/av-2018-11-14b	Gordon, Bell, Kiefel, Keane, Edelman
McKell v The Queen [2019] HCA 5	McKell	http://www.hcourt.gov.au/cases/cases-av/av-2018-12-07	Gordon, Gageler, Kiefel, Nettle, Edelman
OKS v Western Australia [2019] HCA 10	OKS	http://www.hcourt.gov.au/cases/cases-av/av-2019-02-14	Gordon, Gageler, Kiefel, Nettle, Edelman

A.2 An illustration of face landmarking

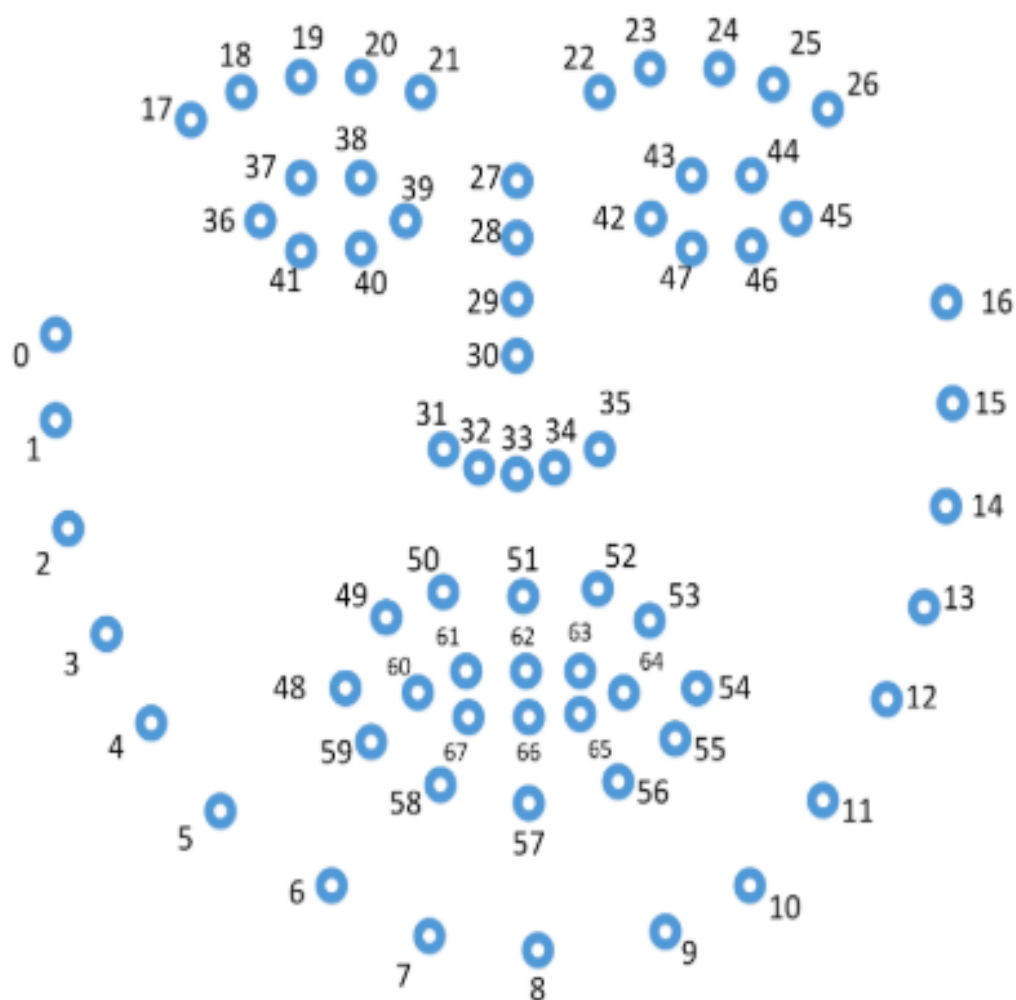


Figure A.1: An illustration of the face landmarking where 67 key points on a face is identified. OpenFace provides 2D and 3D coordinates of these landmarking points

A.3 Description of action units recognised by OpenFace

Table A.2: *The subset of action units OpenFace is able to recognise.*

AU	Meaning	Emotion
AU01	AU01: Inner brow raiser	sadness, surprise and fear
AU02	AU02: Outer brow raiser	surprise, fear and interested
AU04	AU04: Brow lowerer	sadness, fear, anger and confusion
AU05	AU05: Upper lid raiser	surprise, fear, anger and interested
AU06	AU06: Cheek raiser	happiness
AU07	AU07: Lid tightener	fear, anger and confusion
AU09	AU09: Nose wrinkler	disgust
AU10	AU10: Upper lip raiser	NA
AU12	AU12: Lip corner puller	happiness and possibly contempt if appears unilaterally
AU14	AU14: Dimpler	contempt or boredom if appears unilaterally
AU15	AU15: Lip corner depressor	sadness, disgust and confusion
AU17	AU17: Chin raiser	interested and confusion
AU20	AU20: Lip stretcher	fear
AU23	AU23: Lip tightener	anger, confusion or boredom
AU25	AU25: Lips part	NA
AU26	AU26: Jaw drop	surprise and fear
AU28	AU28: Lip suck	NA
AU45	AU45: Blink	NA

A.4 Model estimation result

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Edelman	Nauru-a	AU01	0.6397	0.0294	0.5574	0.715
Nettle	Nauru-a	AU01	0.6955	0.0285	0.6138	0.766
Gageler	Nauru-a	AU01	0.3337	0.0290	0.2606	0.416
Edelman	Nauru-b	AU01	0.5650	0.0364	0.4657	0.659
Nettle	Nauru-b	AU01	0.6734	0.0342	0.5757	0.758
Gageler	Nauru-b	AU01	0.2765	0.0312	0.2009	0.367
Edelman	Rinehart-a	AU01	0.2545	0.0196	0.2054	0.311
Kiefel	Rinehart-a	AU01	0.0589	0.0112	0.0352	0.097
Nettle	Rinehart-a	AU01	0.2167	0.0191	0.1698	0.272
Gageler	Rinehart-a	AU01	0.1295	0.0137	0.0969	0.171
Edelman	Rinehart-b	AU01	0.2567	0.0496	0.1464	0.410

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Nettle	Rinehart-b	AU01	0.3336	0.0567	0.2014	0.498
Gageler	Rinehart-b	AU01	0.2430	0.0473	0.1384	0.391
Edelman	Parkes	AU01	0.1980	0.0229	0.1434	0.267
Keane	Parkes	AU01	0.0625	0.0127	0.0360	0.107
Kiefel	Parkes	AU01	0.0415	0.0092	0.0227	0.075
Bell	Parkes	AU01	0.0196	0.0057	0.0089	0.042
Edelman	McKell	AU01	0.3169	0.0323	0.2370	0.409
Keane	McKell	AU01	0.0672	0.0147	0.0369	0.119
Gageler	McKell	AU01	0.1418	0.0207	0.0946	0.207
Bell	McKell	AU01	0.0423	0.0118	0.0198	0.088
Edelman	OKS	AU01	0.3060	0.0512	0.1873	0.458
Keane	OKS	AU01	0.1652	0.0383	0.0857	0.295
Gageler	OKS	AU01	0.1961	0.0397	0.1103	0.324
Bell	OKS	AU01	0.1898	0.0454	0.0958	0.341
Edelman	Nauru-a	AU04	0.5512	0.0317	0.4654	0.634
Nettle	Nauru-a	AU04	0.2980	0.0304	0.2232	0.385
Gageler	Nauru-a	AU04	0.4886	0.0318	0.4041	0.574
Edelman	Nauru-b	AU04	0.4966	0.0372	0.3979	0.596
Nettle	Nauru-b	AU04	0.2963	0.0345	0.2125	0.396
Gageler	Nauru-b	AU04	0.4447	0.0368	0.3491	0.545
Edelman	Rinehart-a	AU04	0.1167	0.0134	0.0852	0.158
Kiefel	Rinehart-a	AU04	0.0314	0.0076	0.0162	0.060
Nettle	Rinehart-a	AU04	0.0280	0.0049	0.0174	0.045
Gageler	Rinehart-a	AU04	0.1371	0.0150	0.1015	0.183
Edelman	Rinehart-b	AU04	0.2821	0.0529	0.1629	0.442
Nettle	Rinehart-b	AU04	0.1328	0.0333	0.0657	0.250
Gageler	Rinehart-b	AU04	0.5018	0.0623	0.3402	0.663

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Edelman	Parkes	AU04	0.1435	0.0181	0.1012	0.199
Keane	Parkes	AU04	0.1751	0.0214	0.1247	0.240
Kiefel	Parkes	AU04	0.0378	0.0094	0.0192	0.073
Bell	Parkes	AU04	0.2868	0.0265	0.2209	0.363
Edelman	McKell	AU04	0.2056	0.0243	0.1477	0.279
Keane	McKell	AU04	0.1587	0.0231	0.1058	0.231
Gageler	McKell	AU04	0.2027	0.0253	0.1430	0.279
Bell	McKell	AU04	0.4225	0.0360	0.3298	0.521
Edelman	OKS	AU04	0.1188	0.0278	0.0618	0.216
Keane	OKS	AU04	0.2211	0.0428	0.1271	0.356
Gageler	OKS	AU04	0.1706	0.0360	0.0941	0.289
Bell	OKS	AU04	0.6800	0.0518	0.5281	0.801
Edelman	Nauru-a	AU05	0.2579	0.0259	0.1946	0.333
Nettle	Nauru-a	AU05	0.2282	0.0256	0.1665	0.304
Gageler	Nauru-a	AU05	0.5006	0.0315	0.4167	0.585
Edelman	Nauru-b	AU05	0.2324	0.0286	0.1643	0.318
Nettle	Nauru-b	AU05	0.2413	0.0305	0.1687	0.332
Gageler	Nauru-b	AU05	0.4768	0.0373	0.3786	0.577
Edelman	Rinehart-a	AU05	0.2008	0.0175	0.1577	0.252
Kiefel	Rinehart-a	AU05	0.0378	0.0092	0.0195	0.072
Nettle	Rinehart-a	AU05	0.1187	0.0139	0.0861	0.162
Gageler	Rinehart-a	AU05	0.5283	0.0236	0.4648	0.591
Edelman	Rinehart-b	AU05	0.1086	0.0294	0.0511	0.216
Nettle	Rinehart-b	AU05	0.1046	0.0289	0.0484	0.211
Gageler	Rinehart-b	AU05	0.5368	0.0662	0.3616	0.703
Edelman	Parkes	AU05	0.1183	0.0163	0.0811	0.170
Keane	Parkes	AU05	0.4062	0.0292	0.3306	0.486

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Kiefel	Parkes	AU05	0.0197	0.0055	0.0092	0.042
Bell	Parkes	AU05	0.0769	0.0135	0.0476	0.122
Edelman	McKell	AU05	0.1831	0.0221	0.1309	0.250
Keane	McKell	AU05	0.3965	0.0347	0.3078	0.493
Gageler	McKell	AU05	0.4496	0.0342	0.3602	0.542
Bell	McKell	AU05	0.1409	0.0231	0.0894	0.215
Edelman	OKS	AU05	0.0707	0.0179	0.0353	0.136
Keane	OKS	AU05	0.3917	0.0544	0.2584	0.543
Gageler	OKS	AU05	0.3009	0.0495	0.1861	0.448
Bell	OKS	AU05	0.2367	0.0456	0.1359	0.380
Edelman	Nauru-a	AU07	0.3670	0.0300	0.2905	0.451
Nettle	Nauru-a	AU07	0.4918	0.0325	0.4054	0.579
Gageler	Nauru-a	AU07	0.2207	0.0240	0.1630	0.292
Edelman	Nauru-b	AU07	0.3492	0.0347	0.2624	0.447
Nettle	Nauru-b	AU07	0.5251	0.0381	0.4231	0.625
Gageler	Nauru-b	AU07	0.2148	0.0274	0.1502	0.298
Edelman	Rinehart-a	AU07	0.1649	0.0158	0.1266	0.212
Kiefel	Rinehart-a	AU07	0.0514	0.0103	0.0298	0.088
Nettle	Rinehart-a	AU07	0.1720	0.0171	0.1308	0.223
Gageler	Rinehart-a	AU07	0.1298	0.0140	0.0966	0.172
Edelman	Rinehart-b	AU07	0.1621	0.0383	0.0831	0.292
Nettle	Rinehart-b	AU07	0.2669	0.0527	0.1500	0.429
Gageler	Rinehart-b	AU07	0.2376	0.0486	0.1316	0.391
Edelman	Parkes	AU07	0.1346	0.0170	0.0950	0.187
Keane	Parkes	AU07	0.4619	0.0293	0.3846	0.541
Kiefel	Parkes	AU07	0.0393	0.0090	0.0211	0.072
Bell	Parkes	AU07	0.1704	0.0207	0.1216	0.233

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Edelman	McKell	AU07	0.3280	0.0302	0.2525	0.414
Keane	McKell	AU07	0.6076	0.0339	0.5137	0.694
Gageler	McKell	AU07	0.2315	0.0269	0.1670	0.312
Bell	McKell	AU07	0.4314	0.0360	0.3384	0.529
Edelman	OKS	AU07	0.0687	0.0183	0.0331	0.137
Keane	OKS	AU07	0.4036	0.0563	0.2650	0.559
Gageler	OKS	AU07	0.0661	0.0174	0.0321	0.131
Bell	OKS	AU07	0.3902	0.0565	0.2526	0.548
Edelman	Nauru-a	AU14	0.4537	0.0308	0.3729	0.537
Nettle	Nauru-a	AU14	0.4547	0.0322	0.3702	0.542
Gageler	Nauru-a	AU14	0.5287	0.0307	0.4461	0.610
Edelman	Nauru-b	AU14	0.4751	0.0362	0.3798	0.572
Nettle	Nauru-b	AU14	0.5287	0.0375	0.4280	0.627
Gageler	Nauru-b	AU14	0.5606	0.0357	0.4635	0.653
Edelman	Rinehart-a	AU14	0.1624	0.0150	0.1259	0.207
Kiefel	Rinehart-a	AU14	0.1119	0.0148	0.0777	0.158
Nettle	Rinehart-a	AU14	0.1093	0.0129	0.0792	0.149
Gageler	Rinehart-a	AU14	0.2882	0.0206	0.2360	0.347
Edelman	Rinehart-b	AU14	0.1416	0.0348	0.0709	0.263
Nettle	Rinehart-b	AU14	0.1573	0.0380	0.0794	0.288
Gageler	Rinehart-b	AU14	0.4234	0.0627	0.2689	0.595
Edelman	Parkes	AU14	0.2020	0.0212	0.1509	0.265
Keane	Parkes	AU14	0.5648	0.0295	0.4846	0.642
Kiefel	Parkes	AU14	0.1359	0.0187	0.0931	0.194
Bell	Parkes	AU14	0.0092	0.0039	0.0029	0.029
Edelman	McKell	AU14	0.2849	0.0282	0.2155	0.366
Keane	McKell	AU14	0.5401	0.0357	0.4438	0.633

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Gageler	McKell	AU14	0.4046	0.0330	0.3199	0.495
Bell	McKell	AU14	0.0169	0.0072	0.0054	0.052
Edelman	OKS	AU14	0.1477	0.0329	0.0790	0.259
Keane	OKS	AU14	0.5962	0.0564	0.4402	0.735
Gageler	OKS	AU14	0.3148	0.0511	0.1954	0.465
Bell	OKS	AU14	0.0401	0.0176	0.0121	0.125
Edelman	Nauru-a	AU15	0.4242	0.0306	0.3446	0.508
Nettle	Nauru-a	AU15	0.6179	0.0309	0.5322	0.697
Gageler	Nauru-a	AU15	0.4877	0.0309	0.4057	0.570
Edelman	Nauru-b	AU15	0.5538	0.0367	0.4541	0.649
Nettle	Nauru-b	AU15	0.7708	0.0288	0.6844	0.839
Gageler	Nauru-b	AU15	0.6259	0.0350	0.5282	0.714
Edelman	Rinehart-a	AU15	0.1433	0.0142	0.1091	0.186
Kiefel	Rinehart-a	AU15	0.0388	0.0083	0.0216	0.069
Nettle	Rinehart-a	AU15	0.1880	0.0178	0.1447	0.241
Gageler	Rinehart-a	AU15	0.2504	0.0195	0.2016	0.307
Edelman	Rinehart-b	AU15	0.0952	0.0260	0.0446	0.192
Nettle	Rinehart-b	AU15	0.2065	0.0464	0.1083	0.358
Gageler	Rinehart-b	AU15	0.3094	0.0573	0.1788	0.480
Edelman	Parkes	AU15	0.1964	0.0211	0.1456	0.259
Keane	Parkes	AU15	0.7268	0.0254	0.6534	0.790
Kiefel	Parkes	AU15	0.0534	0.0115	0.0297	0.094
Bell	Parkes	AU15	0.2217	0.0237	0.1644	0.292
Edelman	McKell	AU15	0.2236	0.0246	0.1644	0.296
Keane	McKell	AU15	0.6432	0.0331	0.5500	0.727
Gageler	McKell	AU15	0.3196	0.0308	0.2430	0.407
Bell	McKell	AU15	0.2833	0.0313	0.2070	0.375

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Edelman	OKS	AU15	0.0694	0.0177	0.0345	0.135
Keane	OKS	AU15	0.5745	0.0556	0.4226	0.713
Gageler	OKS	AU15	0.1591	0.0338	0.0874	0.272
Bell	OKS	AU15	0.3637	0.0541	0.2335	0.517
Edelman	Nauru-a	AU20	0.7315	0.0256	0.6575	0.795
Nettle	Nauru-a	AU20	0.6862	0.0287	0.6044	0.758
Gageler	Nauru-a	AU20	0.4630	0.0310	0.3814	0.547
Edelman	Nauru-b	AU20	0.7393	0.0298	0.6518	0.811
Nettle	Nauru-b	AU20	0.7376	0.0309	0.6464	0.812
Gageler	Nauru-b	AU20	0.4836	0.0369	0.3861	0.582
Edelman	Rinehart-a	AU20	0.3795	0.0217	0.3231	0.439
Kiefel	Rinehart-a	AU20	0.3075	0.0229	0.2496	0.372
Nettle	Rinehart-a	AU20	0.2363	0.0197	0.1874	0.293
Gageler	Rinehart-a	AU20	0.2302	0.0182	0.1849	0.283
Edelman	Rinehart-b	AU20	0.2815	0.0518	0.1644	0.438
Nettle	Rinehart-b	AU20	0.2617	0.0502	0.1498	0.416
Gageler	Rinehart-b	AU20	0.2901	0.0517	0.1722	0.445
Edelman	Parkes	AU20	0.4594	0.0275	0.3868	0.534
Keane	Parkes	AU20	0.7955	0.0231	0.7265	0.851
Kiefel	Parkes	AU20	0.3712	0.0279	0.2997	0.449
Bell	Parkes	AU20	0.0452	0.0094	0.0256	0.078
Edelman	McKell	AU20	0.6266	0.0310	0.5402	0.706
Keane	McKell	AU20	0.8154	0.0251	0.7383	0.874
Gageler	McKell	AU20	0.4014	0.0326	0.3176	0.491
Bell	McKell	AU20	0.0991	0.0191	0.0583	0.163
Edelman	OKS	AU20	0.5417	0.0558	0.3924	0.684
Keane	OKS	AU20	0.8999	0.0234	0.8170	0.948

Table A.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asympt.LCL	asympt.UCL
Gageler	OKS	AU20	0.4233	0.0542	0.2878	0.571
Bell	OKS	AU20	0.3017	0.0524	0.1813	0.457
Edelman	Nauru-a	AU45	0.2105	0.0260	0.1488	0.289
Nettle	Nauru-a	AU45	0.0576	0.0147	0.0286	0.112
Gageler	Nauru-a	AU45	0.3810	0.0334	0.2961	0.474
Edelman	Nauru-b	AU45	0.1530	0.0254	0.0964	0.234
Nettle	Nauru-b	AU45	0.0486	0.0137	0.0225	0.102
Gageler	Nauru-b	AU45	0.3033	0.0373	0.2130	0.412
Edelman	Rinehart-a	AU45	0.0418	0.0066	0.0274	0.063
Kiefel	Rinehart-a	AU45	0.3845	0.0251	0.3196	0.454
Nettle	Rinehart-a	AU45	0.0063	0.0019	0.0028	0.014
Gageler	Rinehart-a	AU45	0.1348	0.0157	0.0978	0.183
Edelman	Rinehart-b	AU45	0.0333	0.0137	0.0109	0.097
Nettle	Rinehart-b	AU45	0.0088	0.0043	0.0024	0.032
Gageler	Rinehart-b	AU45	0.2078	0.0610	0.0883	0.416
Edelman	Parkes	AU45	0.0812	0.0125	0.0532	0.122
Keane	Parkes	AU45	0.5562	0.0303	0.4739	0.635
Kiefel	Parkes	AU45	0.5475	0.0302	0.4657	0.627
Bell	Parkes	AU45	0.0117	0.0053	0.0034	0.039
Edelman	McKell	AU45	0.0814	0.0139	0.0509	0.128
Keane	McKell	AU45	0.4193	0.0369	0.3245	0.520
Gageler	McKell	AU45	0.2052	0.0271	0.1417	0.288
Bell	McKell	AU45	0.0138	0.0064	0.0040	0.047
Edelman	OKS	AU45	0.0138	0.0052	0.0050	0.038
Keane	OKS	AU45	0.2478	0.0571	0.1262	0.429
Gageler	OKS	AU45	0.0596	0.0199	0.0237	0.142
Bell	OKS	AU45	0.0122	0.0067	0.0028	0.052

A.5 Residual plots

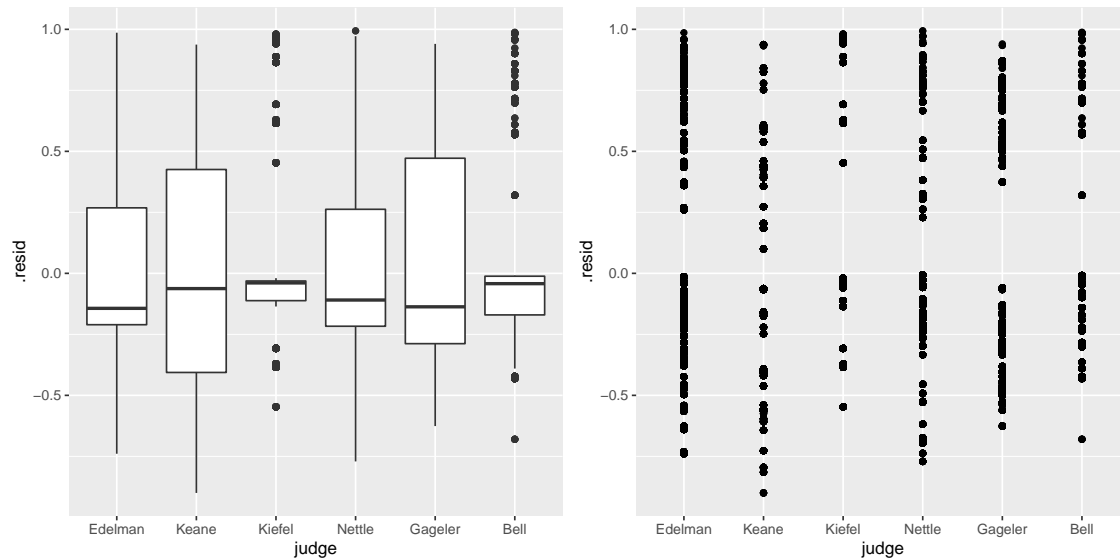


Figure A.2: *Residuals from Model 2 are graphed against judge. On the left panel, the mean of the residuals for each judge is close to zero and on the right panel, there is no clear pattern in the residuals can be observed. This indicates adequate fit.*

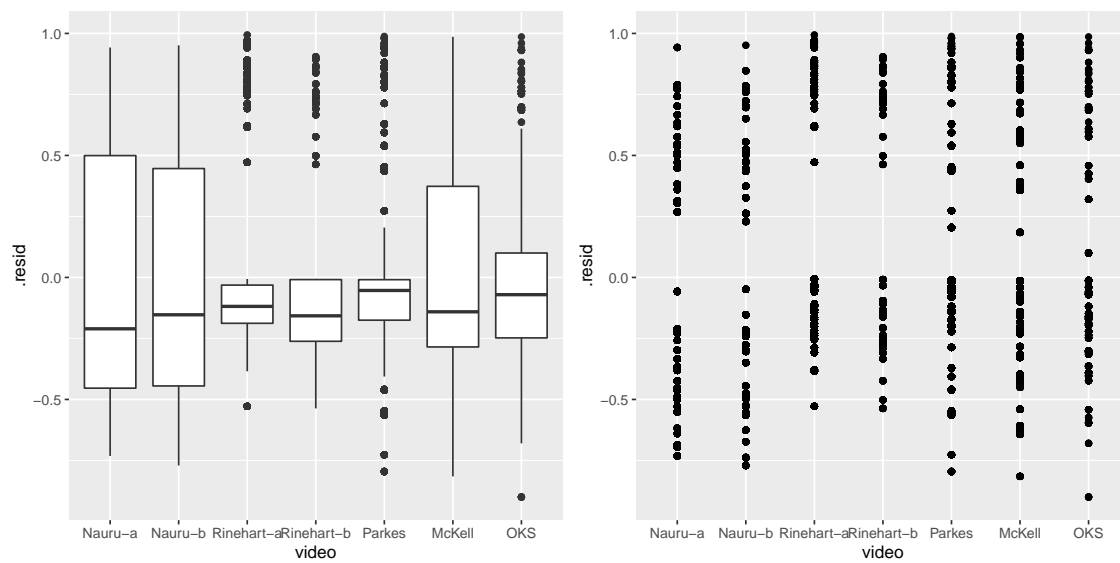


Figure A.3: *Residuals from model 2 are graphed against video. There is no clear pattern in the residuals indicating adequate fit.*

A.6 The presence of facial expression of the justices by speaker

A.7 The intensity of facial expression of the justices by speakers

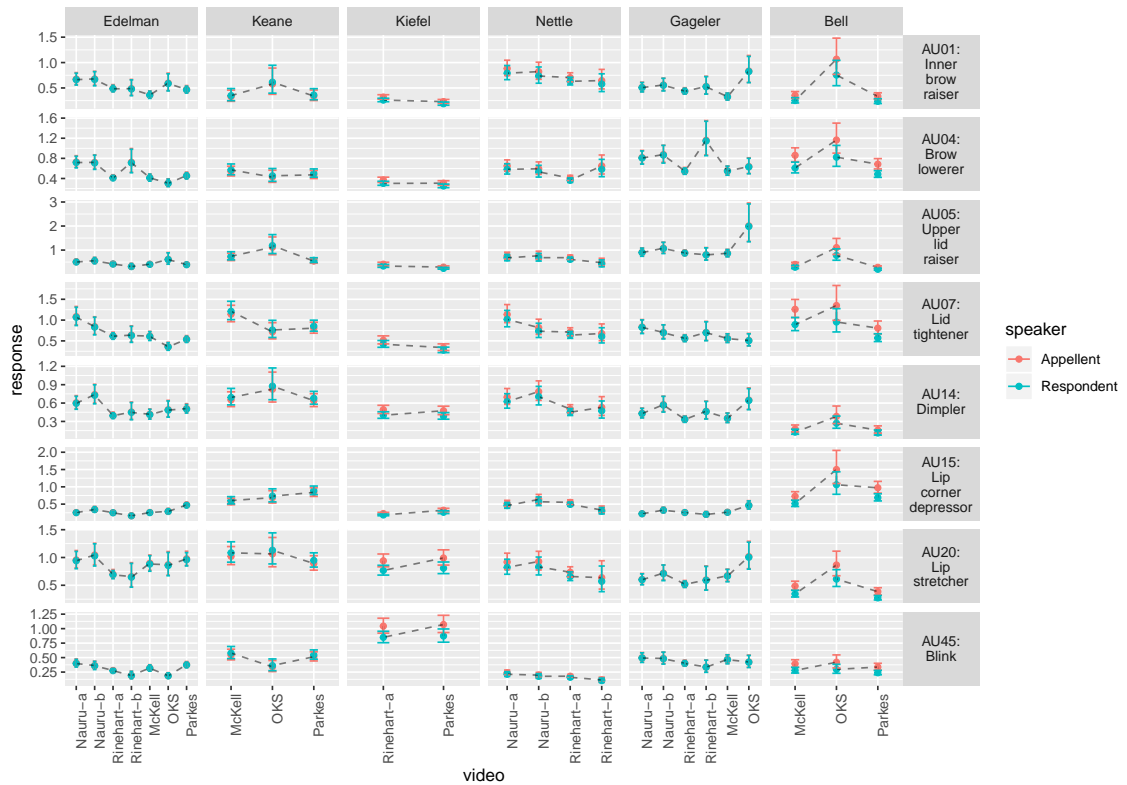


Figure A.4: The 95% confidence interval for estimated mearginal mean of intensity after bonferroni adjustment. The x axis represents video and the y axis represents the estimated marginal mean of the intensity. The facet shows the Justices in columns and action units in rows. The intervals for different speakers are overlaid with different colors.

Bibliography

- Aliotta, J (1987). Combining Judges' Attributes and Case Characteristics: An Alternative Approach to Explaining Supreme Court Decision Making. *Judicature* **71**, 277.
- Baltrusaitis, T, A Zadeh, Y Lim, and L Morency (2018). Openface 2.0: Facial Behavior Analysis Toolkit. In: *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, pp.59–66.
- Black, R, S Treul, T Johnson, and J Goldman (2011). Emotions, Oral Arguments, and Supreme Court Decision Making. *The Journal of Politics* **73**(2), 572–581.
- Chen, D, Y Halberstam, and C Alan (2016). Perceived Masculinity Predicts U.S. Supreme Court Outcomes. *PloS one* **11**(10).
- Chen, D, M Kumar, V Motwani, and P Yeres (2016a). Is Justice Really Blind? And Is It Also Deaf.
- Chen, DL, M Kumar, V Motwani, and P Yeres (2016b). Is Justice Really Blind? And Is It Also Deaf. *And is it Also Deaf*.
- Chief Justices of Australia, TC of and N Zealand (2017). *Guide to Judicial Conduct*. 3rd. Melbourne: Australasian Institute of Judicial Administration Incorporated.
- Cragg, J (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica (pre-1986)* **39**(5), 829.
- Cristy, J, D Lemstra, G Randers-Pehrson, and B Roucres (2019). *ImageMagick*. <https://github.com/ImageMagick>.
- Diehr, P, D Yanez, A Ash, M Hornbrook, and D Lin (1999). Methods for Analyzing Health Care Utilization and Costs. *Annual Review of Public Health* **20**(1), 125–144.
- Dietrich, B, R Enos, and M Sen (2019). Emotional Arousal Predicts Voting on the U.S. Supreme Court. *Political Analysis* **27**(2), 237–243.

- Duan, N, W Manning, C Morris, and J Newhouse (1984). Choosing Between the Sample-Selection Model and the Multi-Part Model. *Journal of Business & Economic Statistics* **2**(3), 283–289.
- Ekman, P, M O’Sullivan, W Friesen, and K Scherer (1991). Face, Voice, and Body in Detecting Deceit. *Journal of Nonverbal Behavior* **15**(2), 125–135.
- Ekman, P and WV Friesen (1976). Measuring Facial Movement. *Environmental Psychology and Nonverbal Behavior* **1**(1), 56–75.
- Epstein, L, W Landes, and R Posner (2010). Inferring the Winning Party in the Supreme Court from the Pattern of Questioning at Oral Argument. *The Journal of Legal Studies* **39**(2), 433–467.
- Faraway, J (2016). *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Chapman and Hall/CRC.
- Gelman, A and J Hill (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge university press.
- Hsuan, Y, R Amine, and M Sergey (2019). Youtube-dl. <https://github.com/ytdl-org/youtube-dl/>.
- Huber, B, D McDuff, C Brockett, M Galley, and B Dolan (2018). Emotional Dialogue Generation Using Image-grounded Language Models. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, pp.277.
- Hyndman, R and Y Khandakar (2008). Automatic Time Series Forecasting: The Forecast Package for R. *Journal of Statistical Software* **26**(3), 1–22.
- Johnson, T, R Black, J Goldman, and S Treul (2009). Inquiring Minds Want To Know: Do Justices Tip Their Hands with Questions at Oral Argument in the U.S. Supreme Court. *Washington University Journal of Law and Policy* **29**, 241.
- Kobakian, S and M O’Hara-Wild (2018). *Taipan: Tool for Annotating Images in Preparation for Analysis*. R package version 0.1.2. <https://CRAN.R-project.org/package=taipan>.
- Kovalchik, S and M Reid (2018). Going Inside the Inner Game: Predicting the Emotions of Professional Tennis Players from Match Broadcasts. In: MIT Sloan Sports Analytics Conference.

- Kulik, C and M Perry E.L.and Pepper (2003). Here Comes the Judge: The Influence of Judge Personal Characteristics on Federal Sexual Harassment Case Outcomes. *Law and Human Behavior* **27**, 69–86.
- Lenth, R (2019). *Emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.4.1. <https://CRAN.R-project.org/package=emmeans>.
- Leung, S and S Yu (1996). On the Choice Between Sample Selection and Two-part Models. *Journal of Econometrics* **72**(1-2), 197–229.
- Liu, L, R Strawderman, M Cowen, and Y Shih (2010). A flexible Two-part Random Effects Model for Correlated Medical Costs. *Journal of Health Economics* **29**(1), 110–123.
- Manning, W, N Duan, and W Rogers (1987). Monte Carlo Evidence on the Choice Between Sample Selection and Two-part Models. *Journal of Econometrics* **35**(1), 59–82.
- Nagel, SS (1962). Testing Relations Between Judicial Characteristics and Judicial Decision-Making. *Western Political Quarterly* **15**(3), 425–437. eprint: <https://doi.org/10.1177/106591296201500301>.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Robinson, D and A Hayes (2019). *Broom: Convert Statistical Analysis Objects into Tidy Tibbles*. R package version 0.5.2. <https://CRAN.R-project.org/package=broom>.
- Schroff, F, D Kalenichenko, and J Philbin (2015). Facenet: A Unified Embedding for Face Recognition and Clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.815–823.
- Shullman, S (2004). The Illusion of Devil’s Advocacy: How the Justices of the Supreme Court Foreshadow Their Decisions During Oral Argument. *Journal of Appellate Practice and Process* **2** **271**(6).
- Steffensmeier, D and C Britt (2001). Judges’ Race and Judicial Decision Making: Do Black Judges Sentence Differently? *Social Science Quarterly* **82**(4), 749–764. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/0038-4941.00057>.
- Taigman, Y, M Yang, M Ranzato, and L Wolf (2014). Deepface: Closing the Gap to Human-level Performance in Face Verification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.1701–1708.

- Tutton, J, K Mack, and S Roach Anleu (2018). Judicial Demeanor: Oral Argument in the High Court of Australia. *Justice System Journal* **39**(3), 273–299.
- Van Koppen, P and J Kate (1984). Individual Differences in Judicial Behavior: Personal Characteristics and Private Law Decision-Making. *Law and Society Review* **18**(2), 225–247.
- Wickham, H (2017). *Tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>.
- Xie, Y (2014). “knitr: A Comprehensive Tool for Reproducible Research in R”. In: *Implementing Reproducible Computational Research*. Ed. by V Stodden, F Leisch, and RD Peng. ISBN 978-1466561595. Chapman and Hall/CRC. <http://www.crcpress.com/product/isbn/9781466561595>.
- Xie, Y (2016). *bookdown: Authoring Books and Technical Documents with R Markdown*. ISBN 978-1138700109. Boca Raton, Florida: Chapman and Hall/CRC. <https://github.com/rstudio/bookdown>.
- Xie, Y, J Allaire, and G Golemund (2018). *R Markdown: The Definitive Guide*. ISBN 9781138359338. Boca Raton, Florida: Chapman and Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.
- Yang, L, D Jiang, L He, E Pei, M Oveneke, and H Sahli (2016). Decision Tree Based Depression Classification from Audio Video and Language Information. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, pp.89–96.