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

I would like to express my gratitude to Professor Di Cook, my supervisors, for detailed guidance and kindness support throughout and Professor Russell Smyth for raising the idea of this project. I would like to appreciate Stephanie Kobakian, with whom I have countless discussion 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, or Judges, restrict emotional displays and 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 seven cases heard in the High Court of Australia. Facial information derived by the software is applied in a logistic model to find the presence of a selected range of action units. The intensity of the action units is modelled by a two part model. It is observed that the 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., 2016) 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 present an observational study of judicial behaviour, based on manually watching the audio footage and taking notes when an obvious emotion is observed. Manually observing the AV recordings may lead to subjective evaluations of facial expression and this motivates us to extend Tutton, Mack, and Roach Anleu (2018)'s work and employ facial recognition technology to study the facial expression of the justices to obtain objective judgements.

1.2 Literature review

The literature summary is divided into two parts: (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. This literature considers characteristics of the Justices and characteristics of the parties in the case; these include gender, political views, religious background of Justices or gender and race of the defendant in criminal cases (Nagel, 1962; Van Koppen and Kate, 1984; Aliotta, 1987; Steffensmeier and Britt, 2001; Kulik and Perry, 2003).

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

Recently, legal studies have focused on the usage of emotion and vocal characteristics of the Justices to predict their decisions. Chief Justices of Australia and Zealand (2017) present 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 value of impartiality discussed by Tutton, Mack, and Roach Anleu (2018) and Goffman (1956). Paul Ekman Ekman et al. (1991) takes a behavioural perspective and suggests that speakers are often unaware of their own facial and vocal inflections. Chen, Halberstam, and Alan (2016); Chen and Halberstam (2017) and Schubert et al. (1992) 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) has used a multilevel logistic model with random effects to suggest that subconscious vocal inflections contain information that is not available from text transcripts. A sizeable study by **chen2018justice** incorporated vocal and image information of the Justices 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. This study showed that image features increased prediction of case outcomes from 64% to 69% and audio features improved prediction of case outcomes from 67% to 69%. This demonstrates the potential of incorporating facial information to understand and predict the decision of the Justices.

The literature often considers the U.S. Supreme Court Database and far less studies have been conducted using Australian High Court data. 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. Tutton's (2018) work has confirmed the potential of using image information to analyse the Justices' behaviour. However this approach could be biased and lead to subjective results influenced by the people observing the videos. Tutton's (2018) study, presents an opportunity to extend image use by utilising facial recognition technology to produce objective results.

1.2.2 Facial recognition

An anatomical study of the decomposition of facial muscles by Ekman and Friesen (1976) led to the development of Facial Action Code (FAC) and identification of the six universal emotions on human faces. This work has laid a solid foundation for analysing facial expression and developing facial recognition software (Huang and Huang, 1997; Tong, Liao, and Ji, 2007).

To analyse facial expressions, effective facial recognition capture technology is needed to extract faces from images. Facial recognition software 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 area in research including depression classification (Yang et al., 2016), emotion studies (Huber et al., 2018) and even 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 decision 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 and was rejected. He then appealed to the Supreme Court of Nauru and won. The appellant, Refugee Status Review Tribunal then appeal to the High Court of Australia. Three High Court Justices sat the hearings were Justices Gageler, Justices Nettle and Justices Edelman and we refer this case as Nauru-a in this project.

Another case from immigration law is TTY167 v Republic of Nauru [2018] HCA 61 is chosen, 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, Justices Nettle and Justices Edelman and we refer to this case as Nauru-b.

Rinehart v Hancock Prospecting Pty Ltd [2019] HCA 13 is a **commercial** case discussing commercial arbitration. The court decided that a non-party to the arbitration agreement can participate in the arbitration by claiming 'through and under' a party to the agreement. This case were held in two hearings due to its complexity and they are 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 the principle of privity of contract while the majority of the Justices interpret the situation broadly.

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 crash of the helicopter by the Parkes city council (respondent). Chief Justices Kiefel, Justices Bell, Keane, Gordon, and Edelman heard the case and it is named Parkes in the project.

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 as 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 and the Court of appeal of the Supreme Court of Western Australia charged the appellant for conviction, the Appellant then appealed to the High Court of Australia. As another criminal case, this time, Justices Bell, Keane, Nettle, Gordon and Edelman also unanimously allowed the appeal and issued a new trial.

A full list of the videos chosen for this study is also available in the Appendix.

Chapter 2

Data Collection

2.1 Data processing

The audio visual recordings of cases described heard by the High Court of Australia (Chief Justices of Australia and Zealand, 2017) are available on the High Court of Australia website. These videos displayed the Justices' faces above the required level of resolution of the OpenFace software, more than 30 pixels.

To analyse the facial expressions of the Justices the videos must be processed by OpenFace. To download videos from the High Court of Australia (Chief Justices of Australia and Zealand, 2017) the software Youtube-dl (Hsuan, Amine, and Sergey, 2019) was used. Image frames were extracted from each of the videos, at every one minute interval via ffmpeg (team, 2019), this resulted in 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) was 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 contained

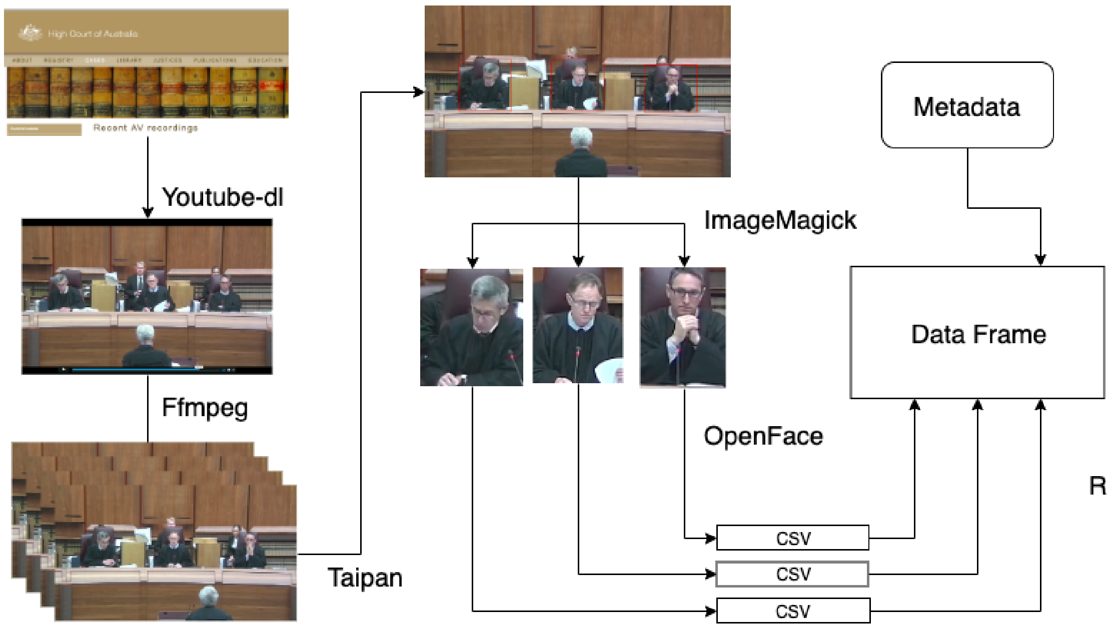

Figure 2.1: *An illustration of the workflow for extracting facial variables from videos.*

facial variables, these include facial landmarking, head pose, eye gaze and action units. The results are stored as separate comma-separated values (csv) files for each of the 4601 faces and post-processing is done in R to combine the separate csv files into a dataframe with additional index columns for frame, judge and video. The workflow to obtain the facial landmarks and expression information from the source videos has been displayed in Figure 2.1.

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. The dataset contains information on the gaze for both eyes with no distinct difference between the eyes.
- **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 6.1 in the Appendix.
- **Action Unit**: An action unit is used to describe the movement of a single facial muscle.

The human facial expression can be de-constructed into a combination of action units. For example, happiness is the addition of action unit 6 and 12, which can be described as a raising check and a pull on the lip corner respectively. The Facial Action Coding System (FACS) is the common standard for describing facial expressions via anatomically decomposing a emotion into action units. To decompose an emotion of sadness, three action units are utilised. Action unit 01 describes the raise of inner brow; action unit 04 describes the general lower of brow and action unit 15 depicts the lower of lip corner. The subset of action units OpenFace is able to recognise is provided in Table 6.2 in the Appendix, along with their meaning and related emotions.

2.3 Data format

The data can be expressed in the long format with action unit as an index and presence and intensity presented as observations in two columns. Table 2.1 presents the data for Justices Edelman in case McKell for all the action units in the first frame in a long format. 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

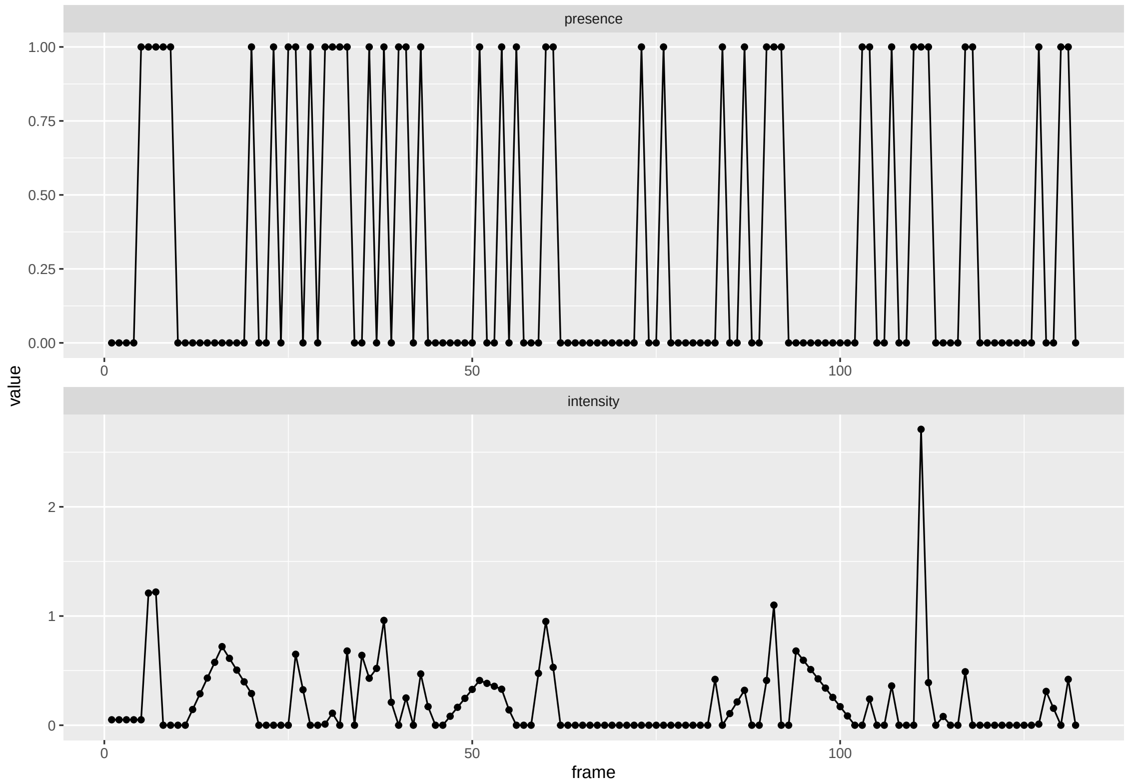

Figure 2.2: *The intensity and presence score of action unit 01 for Justices Edelman in case McKell is graphed against time (frame number) as line chart. The intensity is a numerical variable while presence is binary variable takes value of 0 when the action uit is not present and 1 otherwise.*

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 judges 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. The missing values of intensity are related to missing values in presence. Intensity being zero means the action unit is not present, being one means the action unit is present at minimum intensity and being five means the action unit is present at maximum intensity. Linear interpolation function (`na.interp()`) from `forecast` package is used to impute Intensity. The missing value of presence is then imputed based on if the intensity score of the missing observations are greater than one.

2.5 Source code

Source code for this data pre-processing is available at https://github.com/huizezhang-sherry/ETC4860/data_pre_processing.

Chapter 3

Methodology

3.1 Notation

Let **X** be a matrix of predictors, and **Y** variable in our case is 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 $i = 1, 2, \cdots, 6$
- X_2 indicates `video` for each of the seven cases, $j = 1, 2, \cdots, 7$
- 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, $t = 1, 2, \cdots, T_j$

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_{jitkl} represent the response variable presence, and I_{jitkl} 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. This suggests using a logistic model and we implement this using the `glm()` function from base R. The link function of a matter of choice in the `glm()` function 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 as in Equation 3.1 with the first equation linking the mean of the presence to the linear prediction and the second equation specifying the linkage between η to the predictors. The next section will specify three different function form of the linear predictor by introducing different variables and interactions.

$$\mu = \frac{e^{\eta}}{1 + e^{\eta}} \tag{3.1}$$

$$\eta = f(\alpha_i, \beta_j, \gamma_k, \delta_l) \tag{3.2}$$

3.2.2 Model 1: Action unit

The first linear function is written in Equation 3.3. It includes the main effect of judge and 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.

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

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 judge and action unit variables to differ in different videos, which is useful to answer the research questions *whether the judges are behaving same or different across videos*.

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

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 would be helpful to answer the question *do the expressions of the judges change when different parties are speaking*. The reason for not including more interaction between speaker and video or action unit is because this could cause the model to run out of degree of freedom given the number of observations we have.

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

3.2.5 Model comparison

The analysis of variance (ANOVA) (Faraway, 2016; Gelman and Hill, 2006) is a statistical method that can be used to compare different models. Different packages in R conduct ANOVA test: `anova()` and `drop()` from base R, `Anova()` from `car` package and `aov()`

from `stats` package. We use the base R `anova()` function to compare the three models via chi-square tests.

3.3 Modelling intensity

The intensity score is a continuous variable, with 0 indicating an action unit is not present to a maximum intensity of 5. A histogram of the intensity is plotted in Figure 4.3 and the distribution has a high proportion of zeros with highly skewed continuous value. This type of data is the so-called semi-continuous data (Liu et al., 2010). The semi-continuous data can be modelled in the econometrics literature by the two part model (Cragg, 1971). In the two part model, the data is viewed to be generated via a sequential modelling technique, which is 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 different 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 the problem of not being able to observe Y for those observations with selection variable $z = 0$ doesn't exist in our data. 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 is 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 second model (only one source of zeros).

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

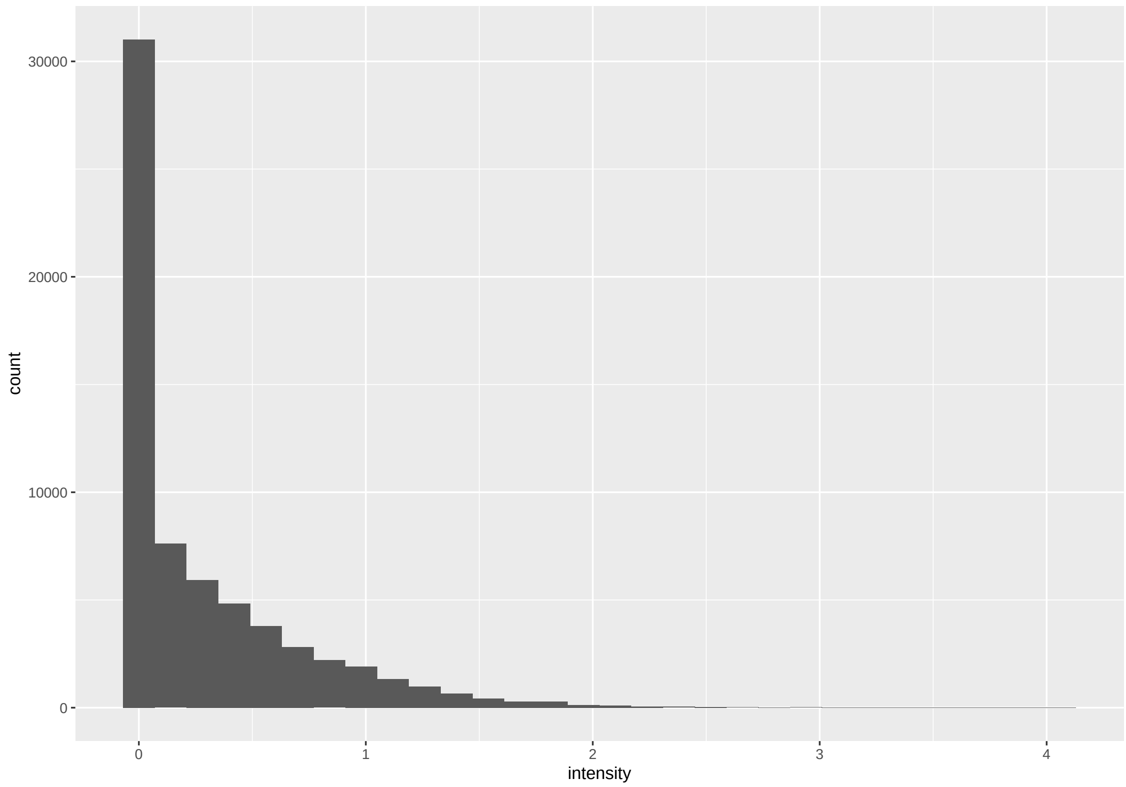

Figure 3.1: *From the histogram of the intensity score, the data is highly skewed with an excessive amount of zeros. The two part model is about to accommodate the excessive zeros via the logistic model and gamma regression is about to capture the skewness in the data.*

$$\mu^1 = \frac{e^\eta}{1+e^\eta} \tag{3.6}$$

$$\eta = f(\alpha_i, \beta_j, \gamma_k, \delta_l) \tag{3.7}$$

$$\mu^2 = \log(I) \tag{3.8}$$

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

The first two equations capture the logistic link and its linear predictor. The next two 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 (**manning1981two**Diehr et al., 1999). More recent literature proposes the use of gamma or generalised gamma regression model for the conditional distribution (Liu et al., 2010). Gamma regression is used to because it could also capture the right skewness and it is easier to implement via the `glm()` function. The log link function is used because the canonical inverse link for gamma distribution will cause some estimated marginal mean 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.

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

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

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

3.4 Post-model analysis

The estimates of variables from the model summary are not particularly useful in our case. This is because firstly, the estimates of the coefficient 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. In our data, the treatment effects include judge, video and action unit. The estimated marginal mean is computed using `emmean()` from the `emmenas` package. The probability from estimated marginal mean have a nice interpretation as the estimated probability of presence score for a particular combination of action unit, judge and video. This output allows us to compare how the estimated presence probabilities of each judge, video and action unit combination 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 (or α in p-value) 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 level so that the type I error of whole family of the simultaneous tests (Family-wise Error Rate (FWER)) is control be less than α. To do this, `confint()` function from base R is used with additional 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 section 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

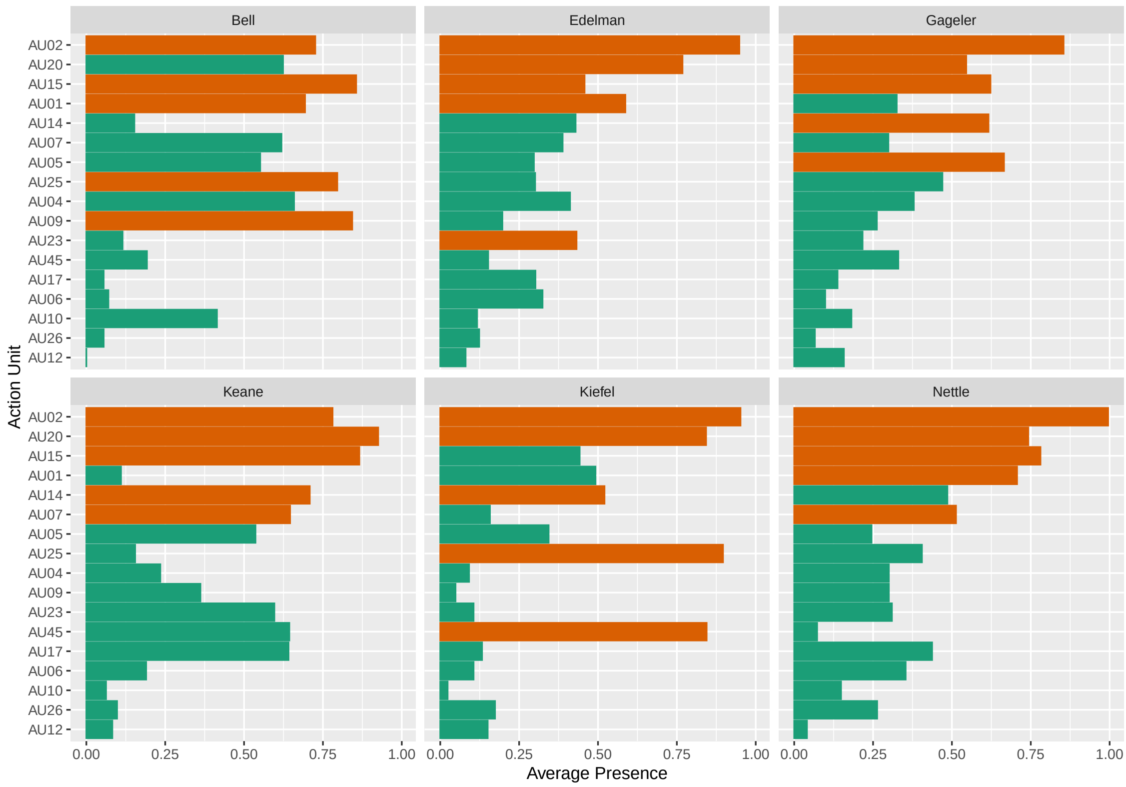

Figure 4.1: *The average presence score of each action unit for each Justice, aggregating on video and time. The most common five action units for each Justices is colored in blue. The most common action units across all the Justices include AU02 (outer eyebrow raise), AU20 (lip stretcher), AU15 (Lip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

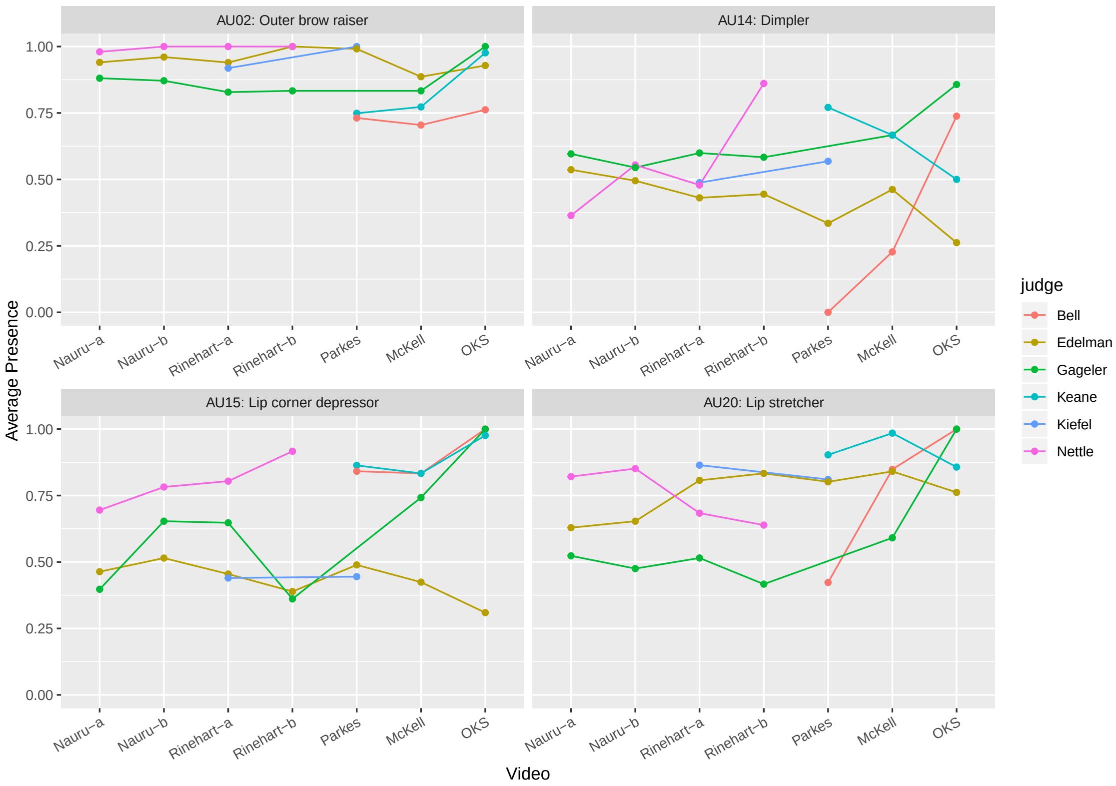

Figure 4.2: *Average presence of the four most common action units for each judge by videos. Some Justices, for example Justices Gageler and Bell show large fluctuation on their facial expressions while others are not.*

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_{ijtk}) 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 through out the elapsed

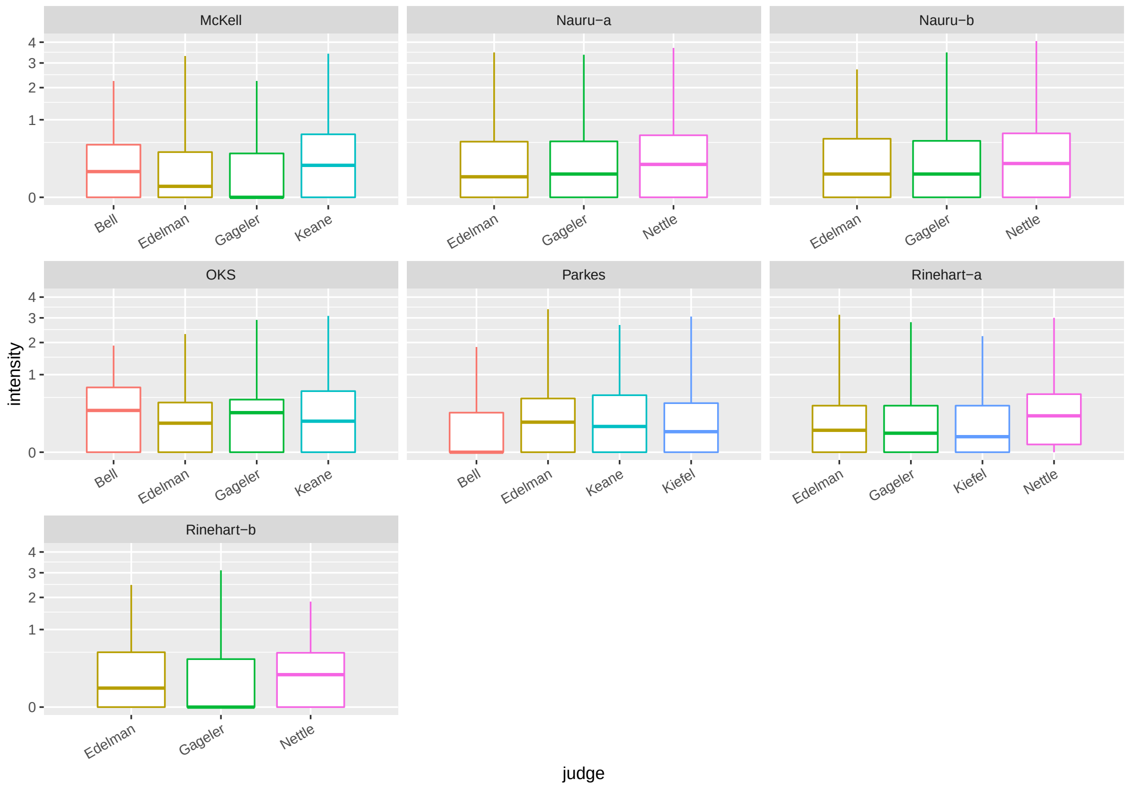

Figure 4.3: *Boxplot of intensity score by Justices and videos. Square root transformation is taken since the mean intensity scores are all below one.*

time in Nauru_a case. The square root transformation is applied so that the mean of the intensity can be easier to visualise. 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 to much of their expressions in the court room. Justices 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 have stronger emotion 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 are more likely to have stronger emotion throughout the time while Justices 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 common 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 seems to have a relatively higher intensity than other Justices.

- Edelman, Gageler and Nettle are the Justices with more intense facial expressions in the courtroom and Justices 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 adn interested
AU07	Lid tightener	fear, anger and confusion
AU14	Dimpler	contempt or boredom if appears unilateraly
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 (0.027) is less than 0.05, however, at 99% significicance 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	30320	35850.07			
Model 2	30259	35255.72	61	594	1.8e-88

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	30259	35255.72			
Model 3	30253	35241.51	6	14	0.027

4.3.2 Residual Diagnostics and post-model analysis

The residuals of Model 2 are plotted against variable judge and video in Figure 6.2 and 6.3 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 6.3 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.

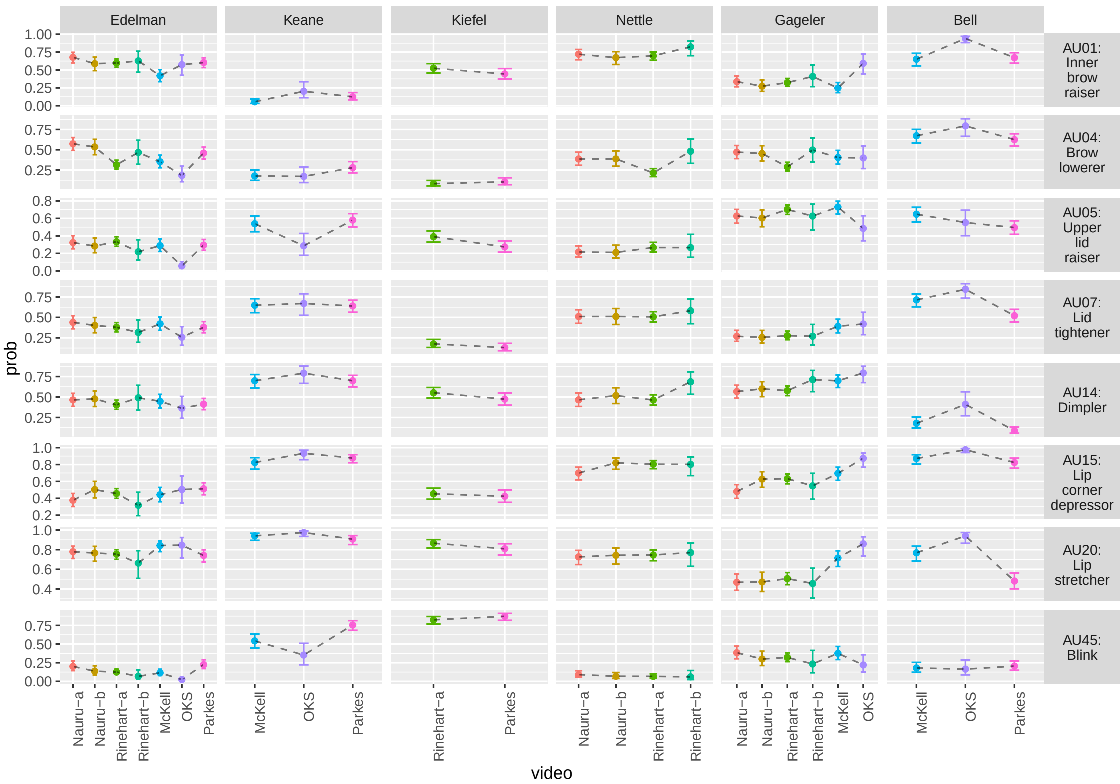

Figure 4.5: *The 95% confidence interval for estimated meargninal mean for presence after bonferroni adjustment. The x axis represents video and the y axis represents the estimated marginal mean of an action unit being observed. The facet shows the Justices in columns and action units in rows.*

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.

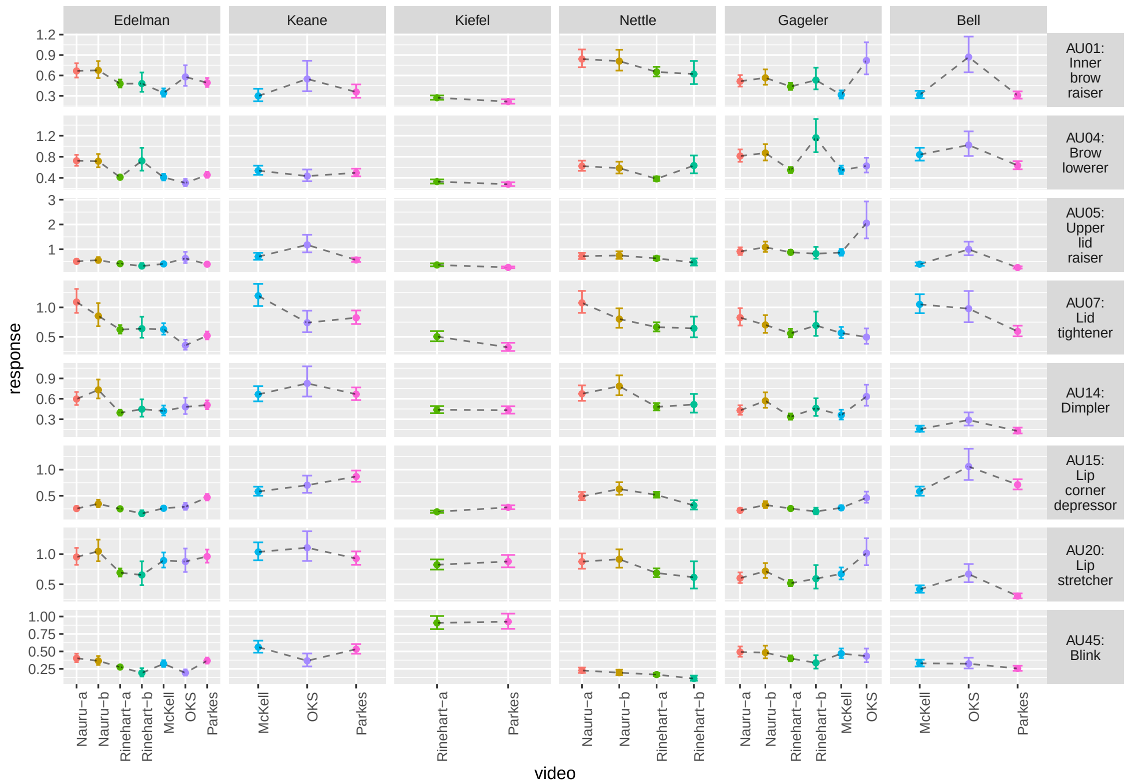

Figure 4.6: *The 95% confidence interval for estimated mearginal mean for presence after bonferroni adjustment. The x axis represents video and the y axis represents the estimated marginal mean of the intensity. The facet shows the Justices in columns and action units in rows.*

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 6.4 and 6.5 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 Grolemund, 2018) and bookdown (Xie, 2016). All matrials required to reproduce the project can be found at `https://github.com/huizezhang-sherry/ETC4860/`.

Chapter 6

Appendix

6.1 List of videos used in the project

Table 6.1: *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

6.2 An illustration of face landmarking

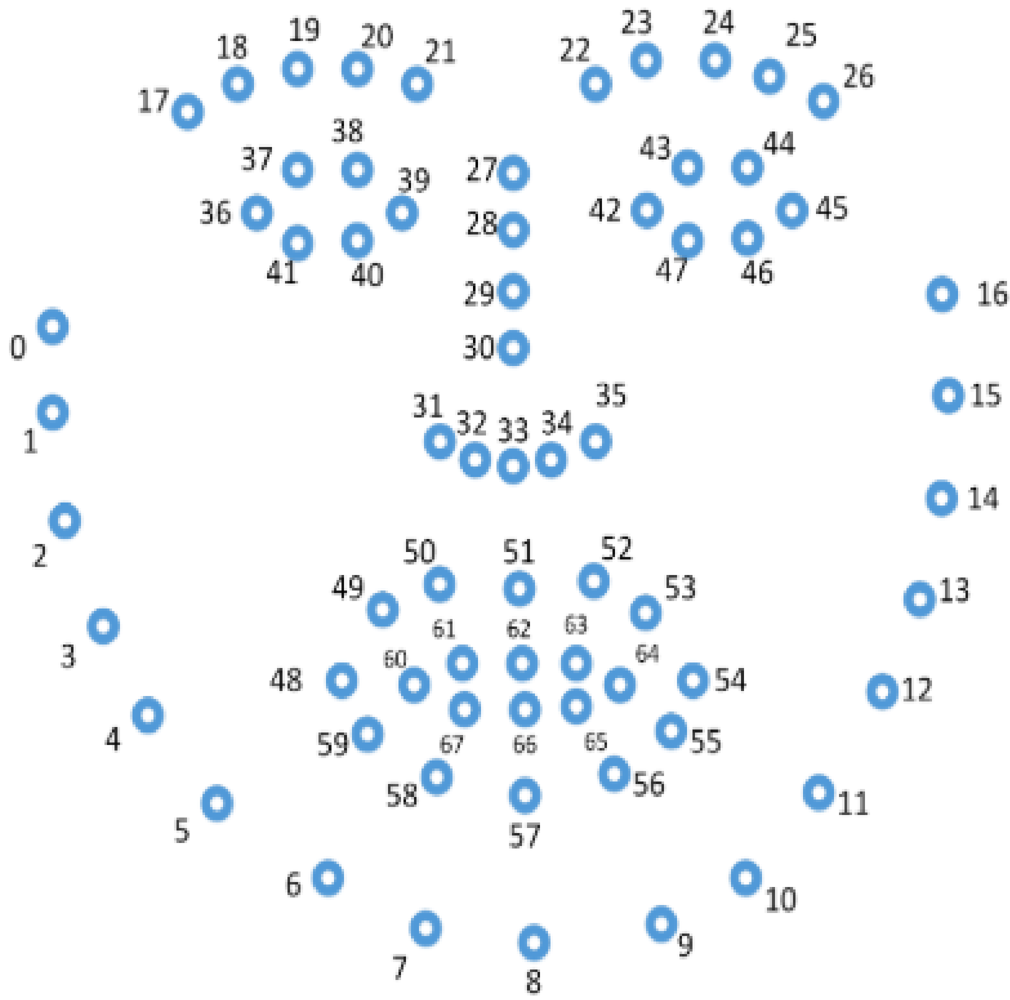

Figure 6.1: *An illustration of the face landmarking where 67 key points on a face is identified. OpenFace provides 2D and 3D coordinates of these landmarking points*

6.3 Description of action units recognised by OpenFace

Table 6.2: *The subset of action units OpenFace is able to recognise.*

AU-number	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 adn 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 unilateraly
AU14	AU14: Dimpler	contempt or boredom if appears unilateraly
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 bordom
AU25	AU25: Lips part	NA
AU26	AU26: Jaw drop	surprise and fear
AU28	AU28: Lip suck	NA
AU45	AU45: Blink	NA

6.4 Model estimation result

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Edelman	Nauru-a	AU01	0.678	0.0276	0.599	0.747
Nettle	Nauru-a	AU01	0.721	0.0269	0.643	0.787
Gageler	Nauru-a	AU01	0.337	0.0283	0.265	0.416
Edelman	Nauru-b	AU01	0.589	0.0354	0.491	0.680
Nettle	Nauru-b	AU01	0.673	0.0337	0.577	0.756
Gageler	Nauru-b	AU01	0.273	0.0302	0.200	0.361
Edelman	Rinehart-a	AU01	0.597	0.0212	0.539	0.652
Kiefel	Rinehart-a	AU01	0.524	0.0245	0.458	0.589
Nettle	Rinehart-a	AU01	0.697	0.0215	0.637	0.752
Gageler	Rinehart-a	AU01	0.323	0.0208	0.270	0.382
Edelman	Rinehart-b	AU01	0.628	0.0561	0.469	0.763

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Nettle	Rinehart-b	AU01	0.824	0.0375	0.700	0.904
Gageler	Rinehart-b	AU01	0.409	0.0577	0.267	0.568
Edelman	McKell	AU01	0.419	0.0316	0.337	0.505
Keane	McKell	AU01	0.055	0.0109	0.032	0.093
Gageler	McKell	AU01	0.247	0.0264	0.183	0.325
Bell	McKell	AU01	0.650	0.0330	0.557	0.733
Edelman	OKS	AU01	0.575	0.0542	0.427	0.710
Keane	OKS	AU01	0.202	0.0414	0.112	0.336
Gageler	OKS	AU01	0.592	0.0533	0.445	0.725
Bell	OKS	AU01	0.941	0.0154	0.883	0.971
Edelman	Parkes	AU01	0.604	0.0256	0.534	0.670
Keane	Parkes	AU01	0.125	0.0187	0.082	0.184
Kiefel	Parkes	AU01	0.445	0.0276	0.372	0.520
Bell	Parkes	AU01	0.672	0.0273	0.595	0.741
Edelman	Nauru-a	AU04	0.573	0.0298	0.492	0.651
Nettle	Nauru-a	AU04	0.386	0.0302	0.309	0.470
Gageler	Nauru-a	AU04	0.471	0.0303	0.391	0.552
Edelman	Nauru-b	AU04	0.535	0.0356	0.439	0.629
Nettle	Nauru-b	AU04	0.387	0.0353	0.297	0.485
Gageler	Nauru-b	AU04	0.454	0.0355	0.361	0.550
Edelman	Rinehart-a	AU04	0.315	0.0202	0.263	0.371
Kiefel	Rinehart-a	AU04	0.082	0.0122	0.055	0.122
Nettle	Rinehart-a	AU04	0.215	0.0184	0.169	0.268
Gageler	Rinehart-a	AU04	0.289	0.0199	0.239	0.346
Edelman	Rinehart-b	AU04	0.467	0.0567	0.322	0.617
Nettle	Rinehart-b	AU04	0.481	0.0578	0.332	0.634
Gageler	Rinehart-b	AU04	0.497	0.0567	0.349	0.645

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Edelman	McKell	AU04	0.352	0.0291	0.278	0.433
Keane	McKell	AU04	0.177	0.0234	0.123	0.249
Gageler	McKell	AU04	0.404	0.0318	0.322	0.492
Bell	McKell	AU04	0.673	0.0315	0.583	0.751
Edelman	OKS	AU04	0.183	0.0354	0.106	0.298
Keane	OKS	AU04	0.171	0.0354	0.096	0.288
Gageler	OKS	AU04	0.399	0.0529	0.268	0.546
Bell	OKS	AU04	0.794	0.0402	0.666	0.882
Edelman	Parkes	AU04	0.458	0.0273	0.386	0.532
Keane	Parkes	AU04	0.279	0.0260	0.215	0.354
Kiefel	Parkes	AU04	0.104	0.0156	0.069	0.154
Bell	Parkes	AU04	0.625	0.0280	0.547	0.697
Edelman	Nauru-a	AU05	0.323	0.0280	0.253	0.402
Nettle	Nauru-a	AU05	0.215	0.0237	0.158	0.286
Gageler	Nauru-a	AU05	0.627	0.0296	0.545	0.703
Edelman	Nauru-b	AU05	0.284	0.0313	0.208	0.375
Nettle	Nauru-b	AU05	0.211	0.0273	0.147	0.294
Gageler	Nauru-b	AU05	0.604	0.0358	0.505	0.695
Edelman	Rinehart-a	AU05	0.333	0.0204	0.280	0.390
Kiefel	Rinehart-a	AU05	0.391	0.0240	0.329	0.457
Nettle	Rinehart-a	AU05	0.267	0.0207	0.215	0.326
Gageler	Rinehart-a	AU05	0.702	0.0202	0.645	0.753
Edelman	Rinehart-b	AU05	0.218	0.0431	0.124	0.355
Nettle	Rinehart-b	AU05	0.266	0.0493	0.155	0.417
Gageler	Rinehart-b	AU05	0.626	0.0567	0.466	0.763
Edelman	McKell	AU05	0.288	0.0269	0.222	0.366
Keane	McKell	AU05	0.539	0.0340	0.447	0.628

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Gageler	McKell	AU05	0.730	0.0272	0.651	0.796
Bell	McKell	AU05	0.647	0.0317	0.558	0.727
Edelman	OKS	AU05	0.057	0.0135	0.029	0.106
Keane	OKS	AU05	0.287	0.0472	0.177	0.428
Gageler	OKS	AU05	0.486	0.0548	0.344	0.630
Bell	OKS	AU05	0.552	0.0558	0.402	0.693
Edelman	Parkes	AU05	0.294	0.0230	0.236	0.359
Keane	Parkes	AU05	0.581	0.0284	0.503	0.655
Kiefel	Parkes	AU05	0.274	0.0240	0.214	0.343
Bell	Parkes	AU05	0.495	0.0289	0.418	0.572
Edelman	Nauru-a	AU07	0.439	0.0301	0.361	0.521
Nettle	Nauru-a	AU07	0.511	0.0312	0.427	0.593
Gageler	Nauru-a	AU07	0.269	0.0253	0.206	0.342
Edelman	Nauru-b	AU07	0.402	0.0350	0.312	0.498
Nettle	Nauru-b	AU07	0.511	0.0368	0.413	0.609
Gageler	Nauru-b	AU07	0.255	0.0290	0.185	0.341
Edelman	Rinehart-a	AU07	0.378	0.0212	0.323	0.437
Kiefel	Rinehart-a	AU07	0.176	0.0182	0.132	0.230
Nettle	Rinehart-a	AU07	0.507	0.0237	0.443	0.570
Gageler	Rinehart-a	AU07	0.276	0.0195	0.227	0.332
Edelman	Rinehart-b	AU07	0.316	0.0517	0.195	0.468
Nettle	Rinehart-b	AU07	0.582	0.0579	0.423	0.725
Gageler	Rinehart-b	AU07	0.269	0.0476	0.161	0.414
Edelman	McKell	AU07	0.421	0.0304	0.342	0.504
Keane	McKell	AU07	0.648	0.0320	0.558	0.729
Gageler	McKell	AU07	0.391	0.0313	0.311	0.478
Bell	McKell	AU07	0.714	0.0294	0.629	0.786

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Edelman	OKS	AU07	0.257	0.0428	0.159	0.387
Keane	OKS	AU07	0.670	0.0499	0.526	0.789
Gageler	OKS	AU07	0.419	0.0522	0.289	0.563
Bell	OKS	AU07	0.844	0.0326	0.735	0.913
Edelman	Parkes	AU07	0.378	0.0257	0.312	0.449
Keane	Parkes	AU07	0.640	0.0275	0.563	0.710
Kiefel	Parkes	AU07	0.131	0.0167	0.092	0.182
Bell	Parkes	AU07	0.521	0.0291	0.443	0.598
Edelman	Nauru-a	AU14	0.464	0.0299	0.386	0.545
Nettle	Nauru-a	AU14	0.465	0.0307	0.384	0.548
Gageler	Nauru-a	AU14	0.567	0.0297	0.486	0.644
Edelman	Nauru-b	AU14	0.477	0.0354	0.384	0.572
Nettle	Nauru-b	AU14	0.517	0.0363	0.420	0.613
Gageler	Nauru-b	AU14	0.600	0.0344	0.505	0.688
Edelman	Rinehart-a	AU14	0.404	0.0210	0.349	0.461
Kiefel	Rinehart-a	AU14	0.552	0.0243	0.486	0.617
Nettle	Rinehart-a	AU14	0.463	0.0235	0.401	0.527
Gageler	Rinehart-a	AU14	0.578	0.0221	0.517	0.636
Edelman	Rinehart-b	AU14	0.491	0.0582	0.340	0.643
Nettle	Rinehart-b	AU14	0.686	0.0519	0.533	0.807
Gageler	Rinehart-b	AU14	0.712	0.0489	0.566	0.825
Edelman	McKell	AU14	0.448	0.0313	0.366	0.533
Keane	McKell	AU14	0.699	0.0308	0.610	0.775
Gageler	McKell	AU14	0.698	0.0286	0.616	0.769
Bell	McKell	AU14	0.178	0.0251	0.121	0.256
Edelman	OKS	AU14	0.363	0.0508	0.240	0.507
Keane	OKS	AU14	0.791	0.0393	0.666	0.878

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Gageler	OKS	AU14	0.793	0.0371	0.676	0.876
Bell	OKS	AU14	0.410	0.0558	0.272	0.564
Edelman	Parkes	AU14	0.412	0.0256	0.345	0.483
Keane	Parkes	AU14	0.698	0.0263	0.623	0.764
Kiefel	Parkes	AU14	0.474	0.0278	0.401	0.549
Bell	Parkes	AU14	0.089	0.0139	0.058	0.134
Edelman	Nauru-a	AU15	0.377	0.0290	0.303	0.458
Nettle	Nauru-a	AU15	0.699	0.0280	0.619	0.769
Gageler	Nauru-a	AU15	0.480	0.0306	0.399	0.562
Edelman	Nauru-b	AU15	0.503	0.0367	0.406	0.601
Nettle	Nauru-b	AU15	0.820	0.0247	0.744	0.877
Gageler	Nauru-b	AU15	0.627	0.0348	0.530	0.715
Edelman	Rinehart-a	AU15	0.457	0.0217	0.399	0.515
Kiefel	Rinehart-a	AU15	0.454	0.0244	0.389	0.520
Nettle	Rinehart-a	AU15	0.804	0.0179	0.751	0.847
Gageler	Rinehart-a	AU15	0.632	0.0217	0.572	0.688
Edelman	Rinehart-b	AU15	0.318	0.0525	0.195	0.472
Nettle	Rinehart-b	AU15	0.802	0.0409	0.669	0.890
Gageler	Rinehart-b	AU15	0.548	0.0588	0.390	0.696
Edelman	McKell	AU15	0.443	0.0318	0.360	0.529
Keane	McKell	AU15	0.823	0.0253	0.745	0.881
Gageler	McKell	AU15	0.696	0.0292	0.612	0.768
Bell	McKell	AU15	0.871	0.0203	0.806	0.917
Edelman	OKS	AU15	0.504	0.0614	0.344	0.663
Keane	OKS	AU15	0.933	0.0194	0.858	0.970
Gageler	OKS	AU15	0.874	0.0301	0.769	0.935
Bell	OKS	AU15	0.975	0.0079	0.942	0.990

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Edelman	Parkes	AU15	0.513	0.0266	0.442	0.584
Keane	Parkes	AU15	0.877	0.0177	0.821	0.917
Kiefel	Parkes	AU15	0.424	0.0275	0.352	0.499
Bell	Parkes	AU15	0.824	0.0220	0.757	0.876
Edelman	Nauru-a	AU20	0.779	0.0233	0.710	0.835
Nettle	Nauru-a	AU20	0.727	0.0269	0.649	0.793
Gageler	Nauru-a	AU20	0.468	0.0311	0.386	0.552
Edelman	Nauru-b	AU20	0.766	0.0282	0.682	0.834
Nettle	Nauru-b	AU20	0.743	0.0305	0.653	0.816
Gageler	Nauru-b	AU20	0.471	0.0369	0.374	0.570
Edelman	Rinehart-a	AU20	0.754	0.0183	0.701	0.800
Kiefel	Rinehart-a	AU20	0.865	0.0157	0.817	0.902
Nettle	Rinehart-a	AU20	0.746	0.0202	0.688	0.796
Gageler	Rinehart-a	AU20	0.506	0.0230	0.444	0.567
Edelman	Rinehart-b	AU20	0.663	0.0536	0.508	0.790
Nettle	Rinehart-b	AU20	0.770	0.0444	0.631	0.868
Gageler	Rinehart-b	AU20	0.455	0.0583	0.308	0.612
Edelman	McKell	AU20	0.842	0.0201	0.780	0.889
Keane	McKell	AU20	0.940	0.0129	0.895	0.967
Gageler	McKell	AU20	0.715	0.0298	0.629	0.788
Bell	McKell	AU20	0.768	0.0283	0.683	0.835
Edelman	OKS	AU20	0.846	0.0381	0.714	0.923
Keane	OKS	AU20	0.974	0.0091	0.934	0.990
Gageler	OKS	AU20	0.859	0.0355	0.735	0.931
Bell	OKS	AU20	0.939	0.0187	0.865	0.974
Edelman	Parkes	AU20	0.741	0.0233	0.674	0.799
Keane	Parkes	AU20	0.907	0.0166	0.852	0.943

Table 6.3: *Estimated marginal means of presence for model 2.*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Kiefel	Parkes	AU20	0.809	0.0215	0.745	0.860
Bell	Parkes	AU20	0.481	0.0301	0.401	0.562
Edelman	Nauru-a	AU45	0.201	0.0238	0.144	0.272
Nettle	Nauru-a	AU45	0.091	0.0157	0.056	0.143
Gageler	Nauru-a	AU45	0.384	0.0319	0.303	0.472
Edelman	Nauru-b	AU45	0.137	0.0224	0.087	0.209
Nettle	Nauru-b	AU45	0.069	0.0144	0.039	0.119
Gageler	Nauru-b	AU45	0.300	0.0362	0.212	0.405
Edelman	Rinehart-a	AU45	0.124	0.0131	0.093	0.164
Kiefel	Rinehart-a	AU45	0.825	0.0185	0.770	0.870
Nettle	Rinehart-a	AU45	0.067	0.0106	0.043	0.102
Gageler	Rinehart-a	AU45	0.321	0.0219	0.265	0.382
Edelman	Rinehart-b	AU45	0.067	0.0216	0.028	0.154
Nettle	Rinehart-b	AU45	0.061	0.0206	0.024	0.146
Gageler	Rinehart-b	AU45	0.233	0.0563	0.115	0.415
Edelman	McKell	AU45	0.114	0.0157	0.078	0.163
Keane	McKell	AU45	0.543	0.0353	0.447	0.635
Gageler	McKell	AU45	0.376	0.0330	0.292	0.468
Bell	McKell	AU45	0.179	0.0244	0.123	0.254
Edelman	OKS	AU45	0.025	0.0069	0.012	0.052
Keane	OKS	AU45	0.353	0.0549	0.222	0.510
Gageler	OKS	AU45	0.220	0.0430	0.126	0.356
Bell	OKS	AU45	0.164	0.0367	0.087	0.288
Edelman	Parkes	AU45	0.225	0.0220	0.171	0.290
Keane	Parkes	AU45	0.756	0.0242	0.685	0.815
Kiefel	Parkes	AU45	0.872	0.0171	0.819	0.911
Bell	Parkes	AU45	0.204	0.0229	0.150	0.273

6.5 Residual plots

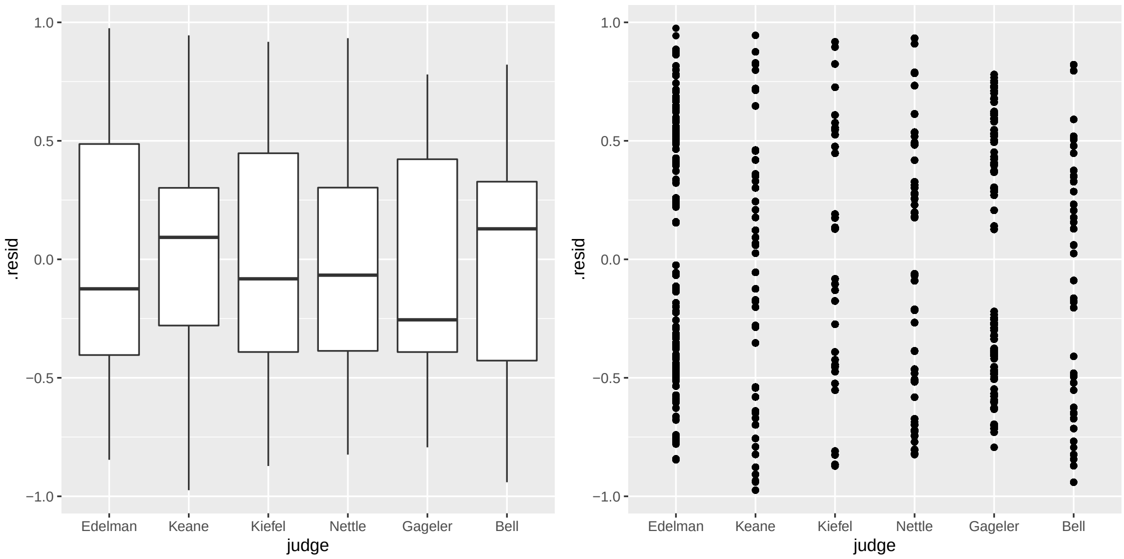

Figure 6.2: *Residuals from Model 2 are graphed against judge. On the left panel, the mean of the residuals for each judge is close to zero and on the right panel, there is no clear pattern in the residuals can be observed. This indicates adequate fit.*

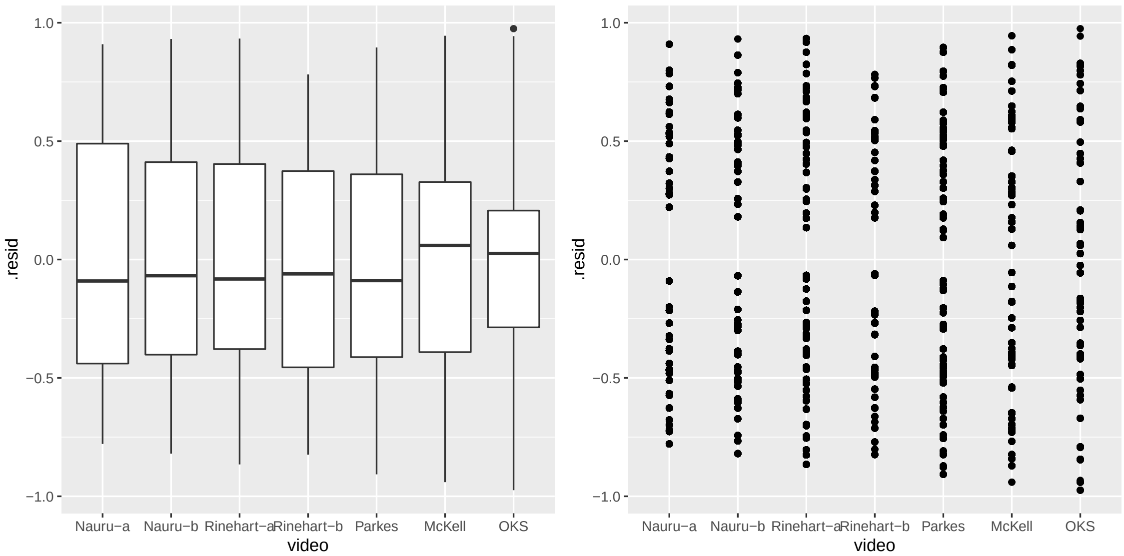

Figure 6.3: *Residuals from model 2 are graphed against video. There is no clear pattern in the residuals indicating adequate fit.*

6.6 The presence of facial expression of the justices by speaker

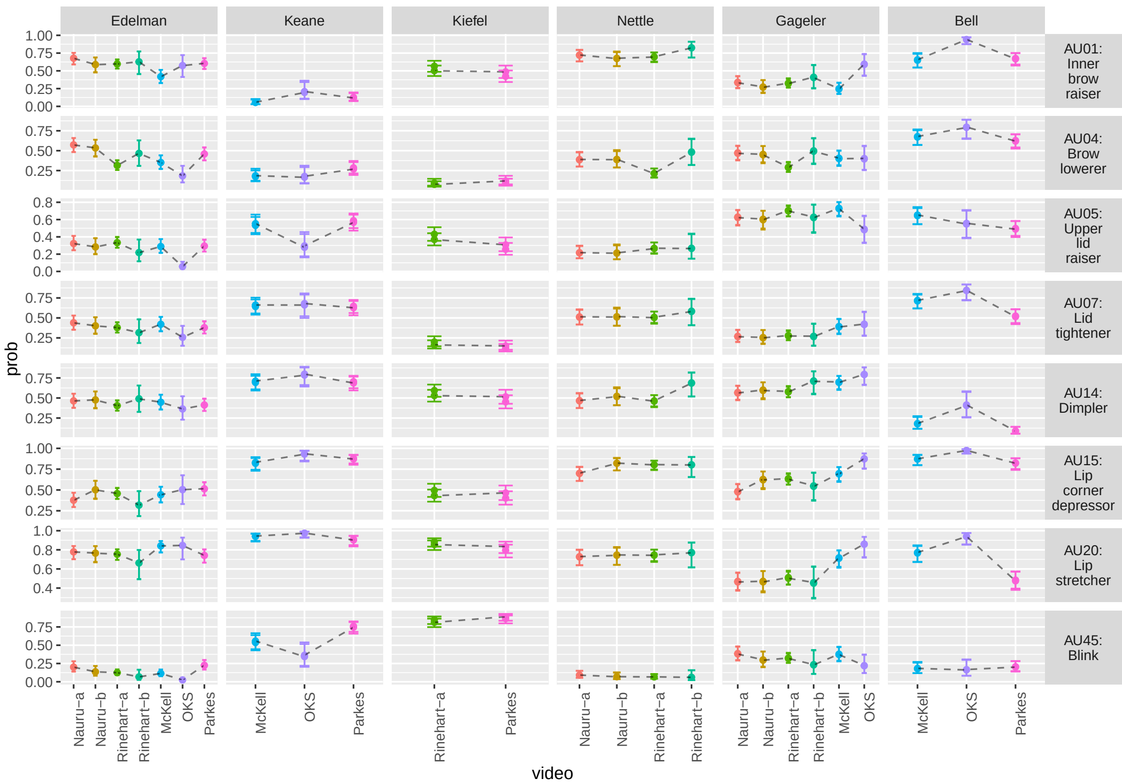

Figure 6.4: *The 95% confidence interval for estimated mearginal mean of presence after bonferroni adjustment. The x axis represents video and the y axis represents the estimated marginal mean of an action unit being observed. The facet shows the Justices in columns and action units in rows. The intervals for different speakers are overlaid with different colors.*

6.7 The intensity of facial expression of the justices by speakers

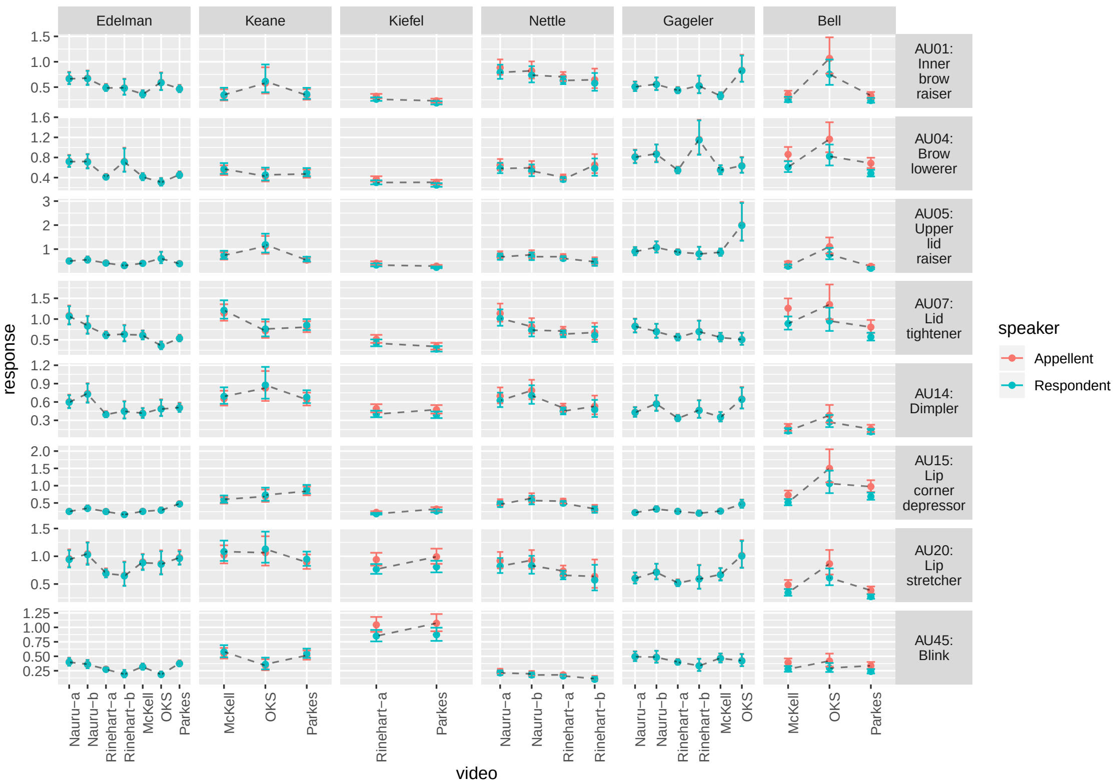

Figure 6.5: *The 95% confidence interval for estimated mearginal mean of intensity after bonferroni adjustment. The x axis represents video and the y axis represents the estimated marginal mean of the intensity. The facet shows the Justices in columns and action units in rows. The intervals for different 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 and Y Halberstam (2017). Covering: Mutable characteristics and perceptions of voice in the U.S. Supreme Court. *Review of Economic Studies invited to resubmit, TSE Working Paper* (16-680).

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 (2016). Is Justice Really Blind? 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 Roucries (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.

Goffman, E (1956). The Nature of Deference and Demeanor. *American Anthropologist* **58**(3), 473–502.

Hsuan, Y, R Amine, and M Sergey (2019). *Youtube-dl*. https://github.com/ytdl-org/youtube-dl/.

Huang, C and Y Huang (1997). Facial Expression Recognition Using Model-Based Feature Extraction and Action Parameters Classification. *Journal of Visual Communication and Image Representation* **8**(3), 278–290.

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.

Schubert, J, S Peterson, G Schubert, and S Wasby (1992). Observing Supreme Court Oral Argument: A Biosocial Approach. *Politics and the Life Sciences* **11**(1), 35–52.

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.

team, F (2019). *ffmpeg*. https://ffmpeg.org/.

Tong, Y, W Liao, and Q Ji (2007). Facial Action Unit Recognition by Exploiting Their Dynamic and Semantic Relationships. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(10), 1683–1699.

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 Grolemund (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.