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

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 the human nature to express emotions as a way to react. However, in particular situation, for example the court, the Justices need to restrict their emotions display as a requirement to ensure the judgement is not biased towards a particular party. In this study, we use facial recognition software to objectively assess the facial expressions of six Justices in seven cases from the high court of Australia. From the obtained facial variables, we model the presence of a selected range of action units by a binomial model and the intensity of the action units by a two part model. From the modelling, we observe that the Justices are remain impartial during the court in general. When a more intense or frequent action unit is presented, it tends to be associated with a negative emotion like sad, fear and anger. Also we find that it would be hard for some Justices to remain a still face in the criminal cases where more extreme behaviour like drug issue and sexual assult are involved.

Chapter 1

Introduction

1.1 Background and motivation

People have attempted to predict the decisions of the Justices in the past century using judge characteristics i.e. Gender, political views, religious background. More recently, scholars(Shullman, 2004; Chen et al., 2018) have been using more information from media(i.e. AV recording, transcript, language used by the Justices) to predict the case outcome using the U.S. Supreme Court data. On-court information has also been used to study data from High Court of Australia. Tutton, Mack, and Roach Anleu (2018) has used an ethnographic approach to present a observational study of judicial behaviour based on watching the audio footage. Manually observing the AV recordings could lead to subjective evaluation of facial expression and this motivates us *to build upon Tutton, Mack, and Roach Anleu (2018)'s work to employ facial recognition technology to study the facial expression of the justices, which will provide a more objective result than Tutton, Mack, and Roach Anleu (2018).*

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 a large law & economics and political science literature that attempts to predict how judges will vote in court cases. Much of this focuses on the characteristics of the judge i.e. gender, political views, religious background and characteristics of the parties in the case i.e. gender or race of the defendant in criminal cases (Nagel, 1962; Koppen and Kate, 1984; Aliotta, 1987-1988; Welch, Combs, and Gruhl, 1988; Steffensmeier and Britt, 2001; Kulik and Perry, 2003).

Moving from static information of the judge and parties involved, more studies start to incorporate the language used by the judge on the court to predict the decision of the Justices. Black et al. (2011) has study 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 with the number of questions asked by the Justices used to infer the winning party in a case.

More recent legal study has focused on the usage of emotion and vocal characteristics of the Justices to predict the judge's votes. Although 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.

and this impartiality has been highlighted in judicial demeanour by Tutton, Mack, and Roach Anleu (2018) and Goffman (1956), Paul Ekman Ekman et al. (1991) suggests that from a behavioural perspective, some facial and vocal inflections are often unbeknown to the speakers themselves. Chen, Halberstam, and Alan (2016); Chen, Halberstam, Yu, et al. (2017) and Schubert et al. (1992) have studied the emotion of the Justices from vocal characteristics and suggest that these vocal characteristics, especially 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.

Moreover, a more sizeable study by Chen et al. (2018) have incorporated both vocal and image information of the judge into a machine learning model to predict the judge votes and case outcome using the U.S. Supreme Court data from 1946-2014. He found that image features improved 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 the decision of the Justices.

The literature mentioned above is mostly conducted using the U.S. Supreme Court Database and less studies have been conducted using Australian High Court data. Tutton, Mack, and Roach Anleu (2018) has used an ethnographic approach to study the judicial demeanour in the High Court of Australia and it is the first of its kind to use transcript and AV recordings in Australian study. The study found that Justices present a detached facial demeanour during the court in most of the time while some human display of emotions i.e. laughter and humour have also been captured by the scholars. Tutton's work has confirmed the potential of using image information to understanding the Justices as in Chen's study, while the ethnographic approach could be biased and lead to subjective results when different people are observing the videos. Thus, building upon Tutton's study, my work fills the gap of producing objective result via utilising facial recognition technology.

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) (Ekman and riesen, 1978) 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 for researchers (Kobayashi and Hara, 1992; Huang and Huang, 1997; Lien et al., 2000; Kapoor, Qi, and Picard, 2003; Tong, Liao, and Ji, 2007; Cohn et al., 2009; Lucey et al., 2010).

To be able to analysis the facial expression, proper facial recognition technique is needed to first extract faces from images. Facial recognition software i.e. DeepFace (Taigman

et al., 2014) from Facebook and FaceNet (Schroff, Kalenichenko, and Philbin, 2015) from Google have also been developed for face detection. 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; Nasir et al., 2016b), emotion study (Pan and Hamilton, 2018; Nasir et al., 2016a; Huber et al., 2018) and even sports analytics. (Kovalchik and Reid, 2018).

Chapter 2

Data Collection

2.1 Data Processing

The source data of this research project is the AV recordings publicly available from the High Court of Australia (Australia, 2019). Due to the requirement of resolution (more than 30px for face detection) of OpenFace, we picked up seven cases from 2018 that have less than seven judges as the sample videos for our dataset. A full list of video being processed can be found in Table **??** in the Appendix.

Multiple procedures need to be performed to obtain the numerical value of facial variables from the source videos. The entire workflow has been plotted in Figure 2.1. Youtube-dl (Hsuan, Amine, and Sergey, 2019) has been used to download videos from the High Court of Australia (Australia, 2019). Image frames are extracted from the videos for every minute via ffmpeg (Bellard, 2019), resulting in 1021 image frames. Because the Justices are remain seated at the same position throughout the hearing, Taipan (Kobakian and O'Hara-Wild, 2018) is used to find the x-y coordinates of the location of the Justices in each image frame. ImageMagick (Cristy et al., 2019) is followed to crop the face of each Justice from each image frame based on the coordinates from Taipan. The resulting 4601 cropped faces are then sent to OpenFace (Baltrusaitis et al., 2018) to process on Docker. The output of OpenFace contains facial variables like facial landmarking, head pose, eye gaze and action unit. These outputs are stored as separate comma-separated values (csv)
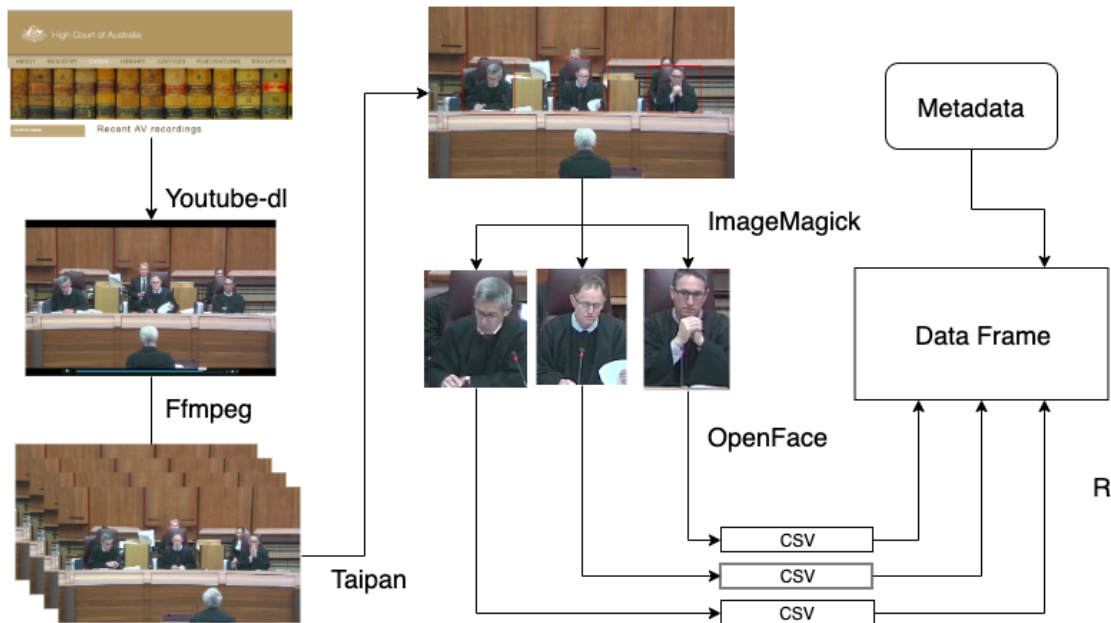
Figure 2.1: *This is an illustration of the workflow for extracting facial variables from videos.*

files for each of the 4601 faces and post-processing is done in R to combine all the separate csv files into a final dataframe with appropriate index of frame, judge and video added.

2.2 Variable description

OpenFace provides more than 711 variables measuring different aspects of a given face and 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 confidence OpenFace is with the detection. Confidence is related to the angle that the Justice's face present in the images.

- **Gaze**: Gaze tracking: the vector from the pupil to corneal reflection. The dataset contains information on the gaze for both eyes while there is no distinct difference between the eyes. Also I was trying to make animation to track the change of the gaze for judges but no good luck.

- **Pose**: the location of the head with respect to camera. Pose-related variables don't provide much useful information apart from gaze-related variables.

- **Landmarking**: landmarking variables for face and eyes. Landmarking variables allows me to plot the face of the judge in a particular frame. More work could be done to explore the usefulness of landmarking variables.

- **Action Unit**: Action units are used to describe facial expressions. The action unit has intensity measures ending with _c and presence measures ending with _r.

In this project, we will focus on using action units along with information about judge, video, speaker to analyse the face of the judge.

2.3 Data format

The data can be expressed in the long format with action unit being an index and presence and intensity being two columns. The Table 2.1 presents the data for Justices Edelman in video McKell for all the action unit in the first frame in long format. Figure 2.2 plots the presence and intensity across time.

Table 2.1: *This table is an illustration of a proportion of the data 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
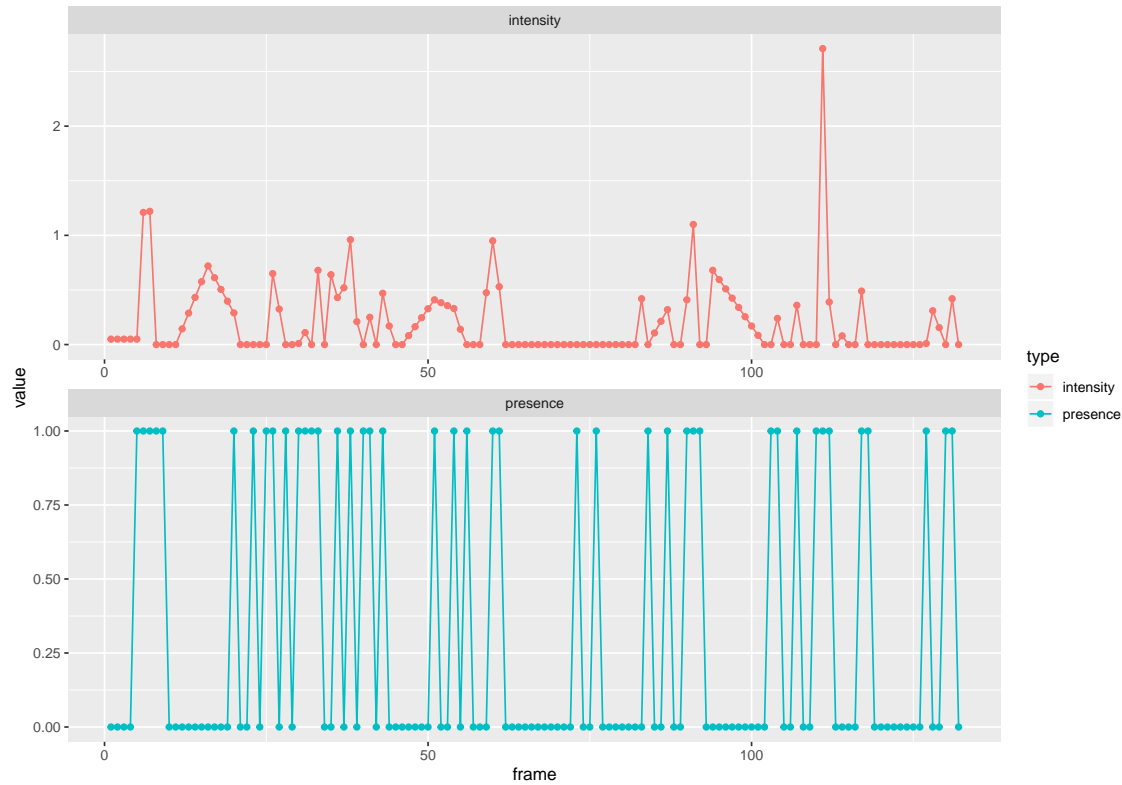
Figure 2.2: *The plot shows the intensity and presence score of action unit 01 for Justices Edelman in case McKell in a time series manner.*

2.4 Missing value imputation

The missingness in the dataset could be due to the fact that a judge is reading the materials on the desk so the face is not captured for a particular frame or simply because some faces are not detectable for the given resolution of the video stream. Simply drop the missing observation will cause the time interval to be irregular and thus imputation is needed.

Intensity is a continuous variable that measures how strong the action unit is presented. It is zero when the action unit is not present, 1(present at minimum intensity) to 5(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

The whole data processing procedure is done as a major part of the project and one of the difficulties of this project is actually to source reliable facial data of the judges, so that the subsequent modelling can then be conducted meaningfully. The whole workflow has been scripted and labelled accordingly from `data_01...` to `data_07...` on my github repository. If re-processing is needed, or more videos need to be added, it is straightforward for others to do.

Chapter 3

Methodology

3.1 Notation

Let $\mathbf{X}$ be a matrix of predictors, and $\mathbf{Y}$ 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 theoratically, 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 Analysis of variance (ANOVA)

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 starts from zero when the action unit is not present to 5, where an action unit is presented at maximum intensity. 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 (Neelon and O'Malley, 2019; Liu et al., 2010). The semi-continuous data can be modelled in the econometrics literature by the two part model(Cragg, 1971; Manning et al., 1981). 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 $y \mid y > 0$.
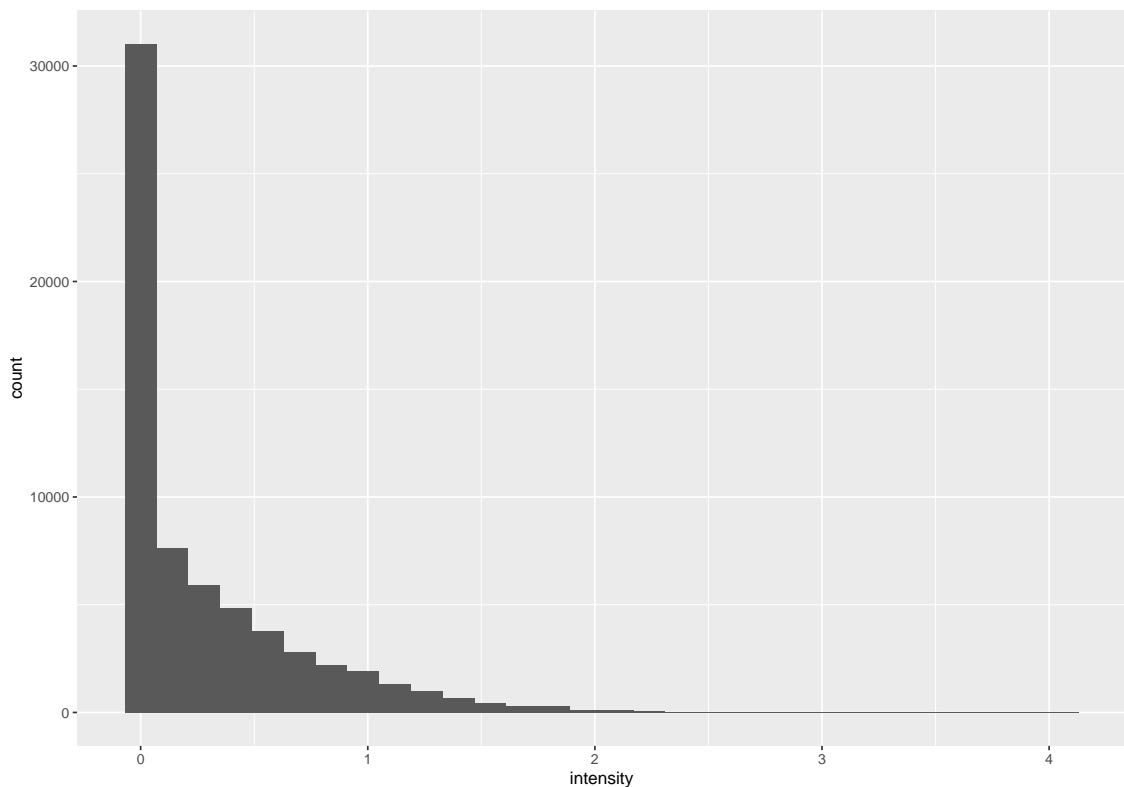
Figure 3.1: *From the histogram of the intensity score, the data is highly skewed with an excessive amount of zeros. The two part model is about to accommodate the excessive zeros via the logistic model and gamma regression is about to capture the skewness in the data.*

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.

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

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

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

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

The first two equations capture the logistic link and its linear predictor. The next two specify the functional form of the conditonal 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 (Manning et al., 1981; 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 treatmetn 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

Testing significance based on p-value has been long criticised for its interpretation. Researchers can erroneously conclude significance because of p-value being less than 0.05 without discussing the false positive/negative proportion. On the other hand, confidence interval provides a confidence range for the estimates to highlight the uncertainty around estimation. Thus confidence interval is used to compare whether the estimated mean for a particular judge-AU group is same or different across videos based on if the intervals overlap with each other.

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

The average presence (P_{ik}) of each action unit is first computed for each judge as

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

This is plotted in Figure 4.1 to give an overview of 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 for each judge and summarised in Table 4.1

Table 4.1: *The five most commonly presented action units for each judge.*

index	Bell	Edelman	Gageler	Keane	Kiefel	Nettle
1	AU15	AU02	AU02	AU20	AU02	AU02
2	AU09	AU20	AU05	AU15	AU25	AU15
3	AU25	AU01	AU15	AU02	AU45	AU20
4	AU02	AU15	AU14	AU14	AU20	AU01
5	AU01	AU23	AU20	AU07	AU14	AU07
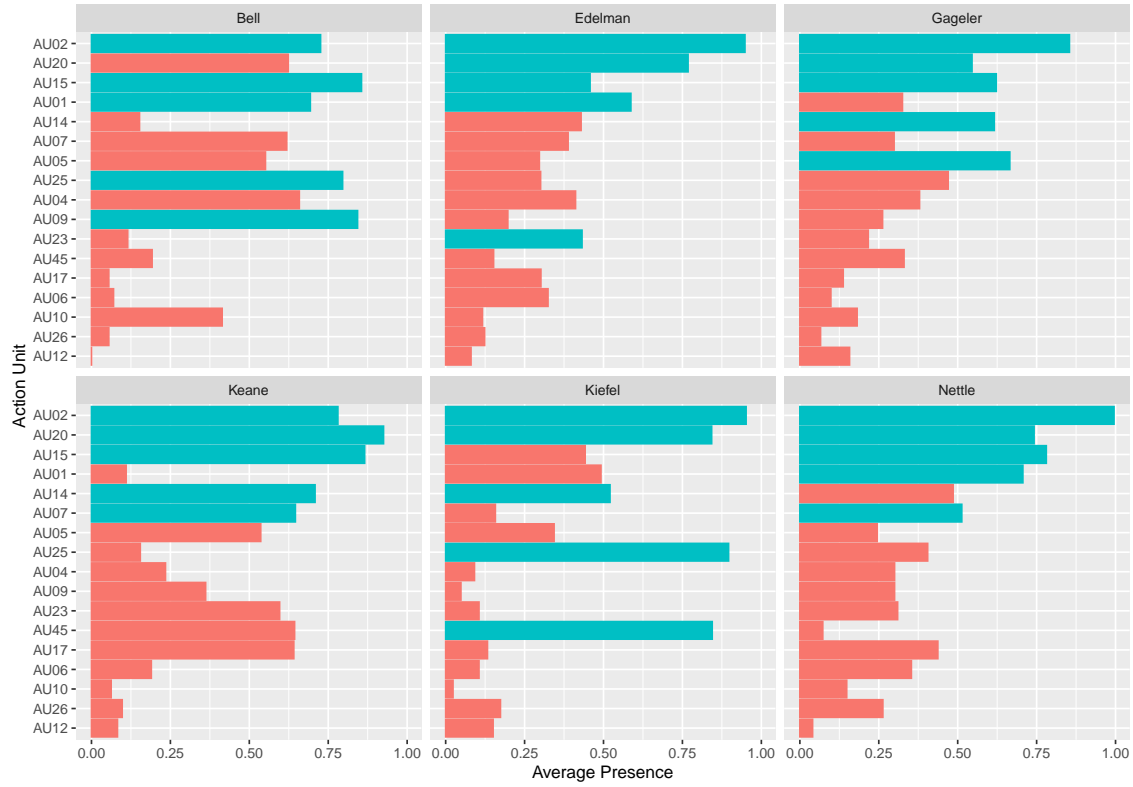
Figure 4.1: *The average presence score of each action unit for each judge, aggregating on video and time.*

It can be seen that some of the action units are common across almost all the judges, these includes AU02 (outer eyebrow raise), AU20 (lip stretcher), AU15 (Lip Corner Depressor) and AU14 (Dimpler).

According to Ekman, Friesen, and Hager (2002), AU02 makes a contribution to surprise, which may be a positive attitude showing that judges are interested in a particular moment. AU14 indicates boredom and AU15 shows confusion. Based on the most common five action units, the emotions judges displayed in the courtroom can be summarised into three categories, described in Table **??**, along with the featured action units.

Presence by videos

We are also interested in the main presence score of the judges by video (P_{ijk}). This is computed as

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

22
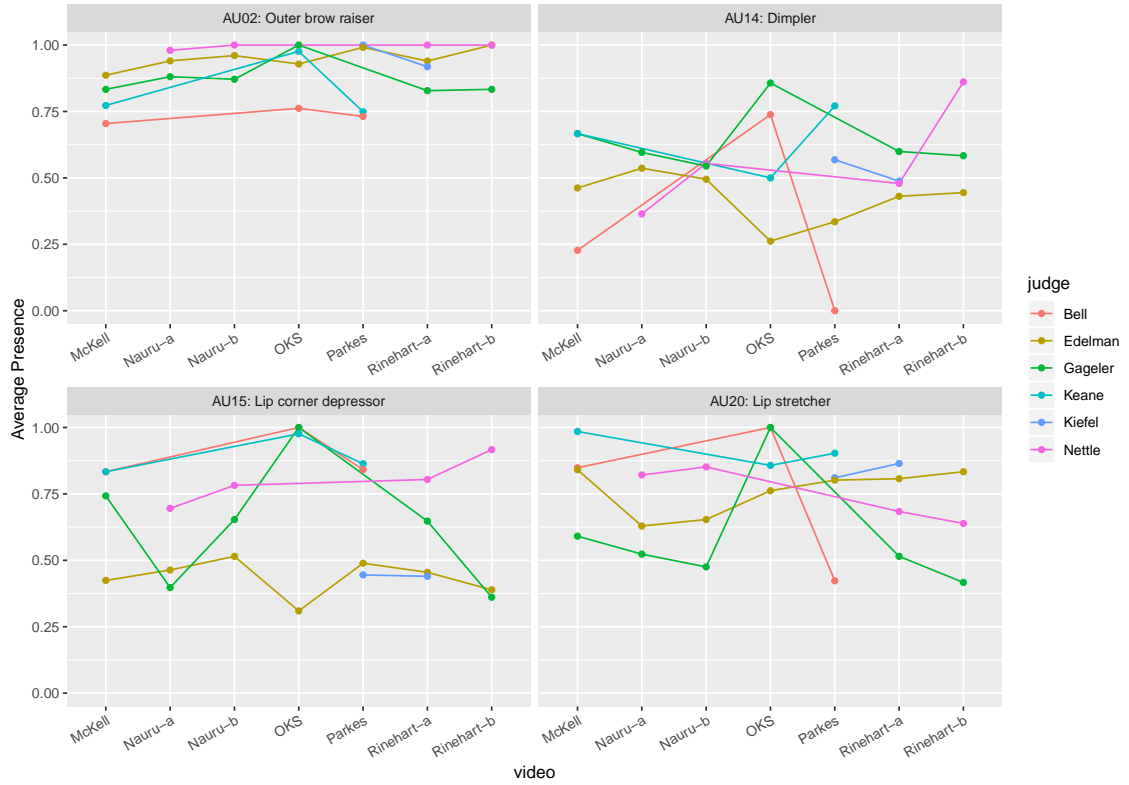
Figure 4.2: *Average presence of the four most common action units for each judge by video*

for the four most common action units: AU02, AU14, AU15, AU20 and plotted in Figure 4.2. From this plot, we can observe that some Justices have larger fluctuation of their expression of action units while others are more consistent throughout different videos. For example, Justice Gageler, coloured as green, has a much higher proportion of expression in case OKS, especially in action unit 14, 15 and 20. Justices Bell, who is coloured red has the highest proportion of expression in case OKS while lowest in case Parkes for action unit 14 and 20.

4.1.2 Action unit: Intensity

General Intensity plot

The boxplot of the intensity for all the judges 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 judge i in a specific case j. For example, the first bar-and-whisker in case Nauru_a is created using all the 17 action units of Edelman through out the elapsed time in Nauru_a case. The square root transformation is applied so that the mean of the
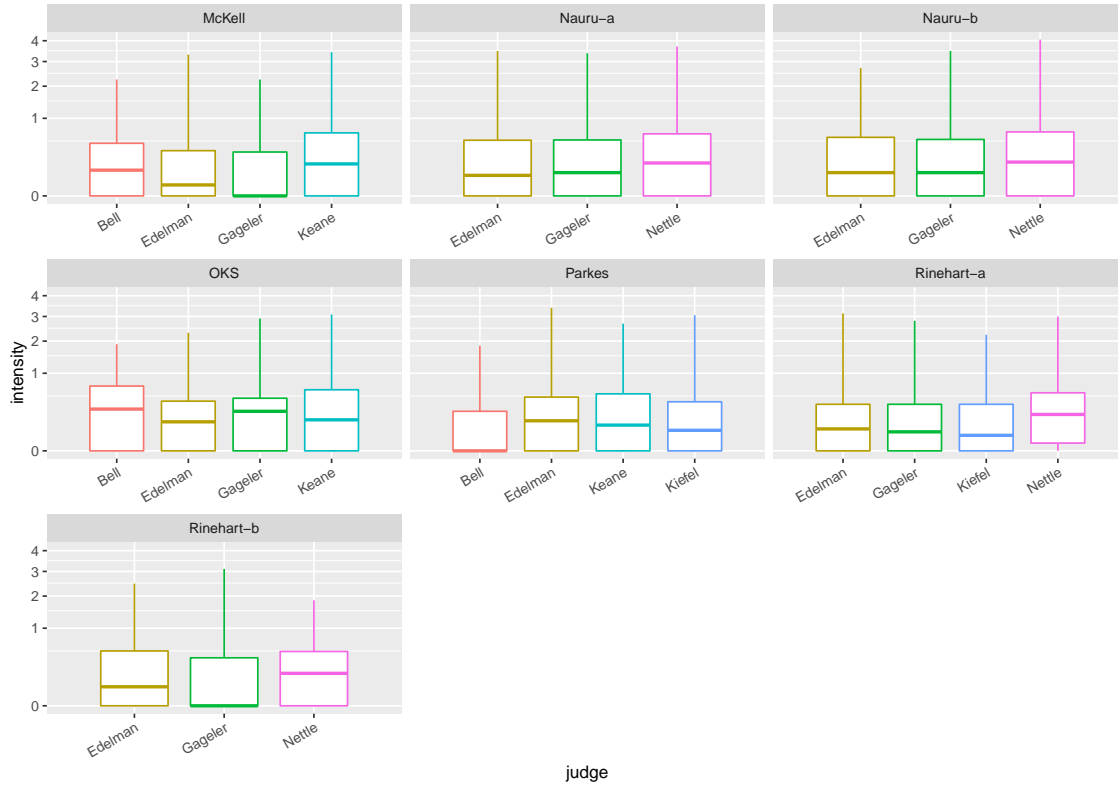
Figure 4.3: *General intensity score by judge and video*

intensity can be easier to be visualised. From the plot, we can see that most of the action

units have low intensity score and this is expected because usually judges are expected

not to express to much of their expressions in the court room. We can find that Justices

Nettle, colored pink seems to have higher average in all the four cases he appeared.

Mean intensity

Mean intensity score (I_{ik}) of each action unit for each of the judge is computed as

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

and plotted in Figure 4.4. The five most intense action units for each judge are presented

in Table 4.2. We can find that the most common high intense action units includes AU20

(Lip Stretcher), AU07 (Lid Tightener) and AU04 (Brow Lower).
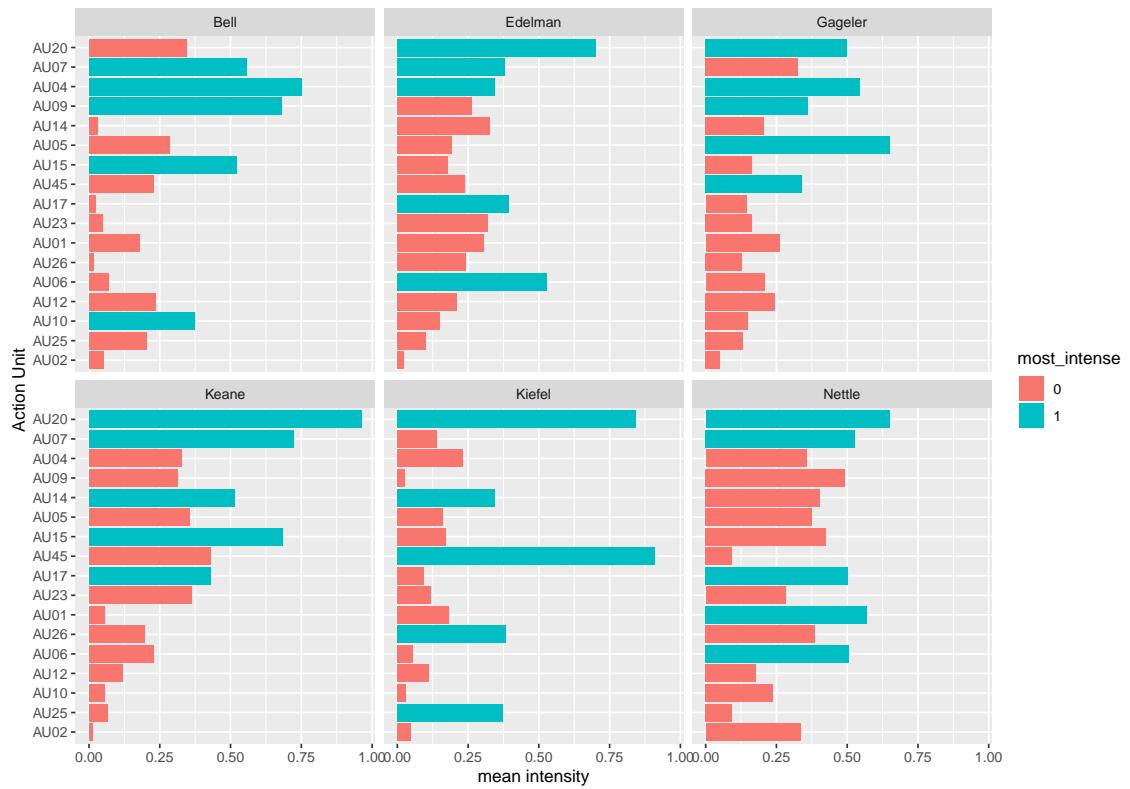
Figure 4.4: *Mean intensity score for each judge and action unit aggregating on videos.*

Table 4.2: *The five most intense action unit for each judge.*

index	Bell	Edelman	Gageler	Keane	Kiefel	Nettle
1	AU04	AU20	AU05	AU20	AU45	AU20
2	AU09	AU06	AU04	AU07	AU20	AU01
3	AU07	AU17	AU20	AU15	AU26	AU07
4	AU15	AU07	AU09	AU14	AU25	AU06
5	AU10	AU04	AU45	AU17	AU14	AU17

High intensity points

I also plotted the points with intensity greater than two against time for all the justices in Figure 4.5. It tells us that Edelman, Gageler and Nettle are the judges have stronger emotion that can be detected since they have more points with intensity greater than 2. Different judges 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.
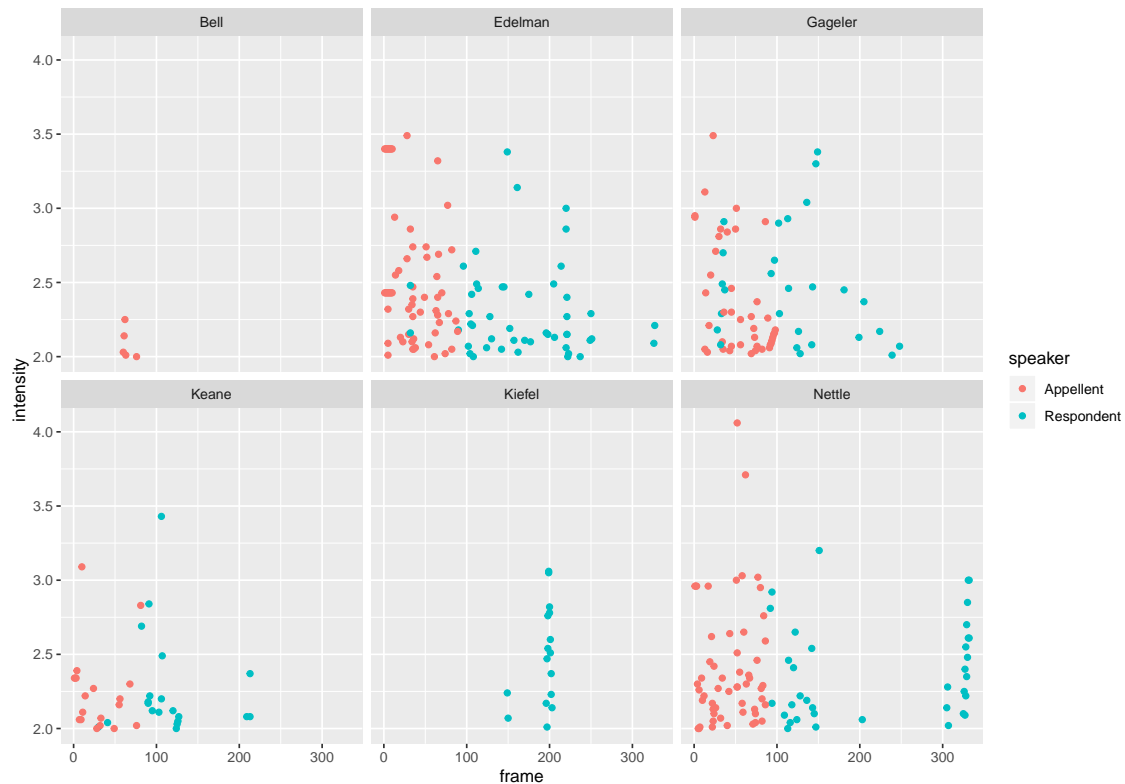
Figure 4.5: *this is the plot*

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 raise), 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 Choice of action unit to include

The number of action unit to include in the model is a matter of choice. The discussion of this choice is to ensure our 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 for both mean presence and intensity rank. This ensures that these action units have both relativcely high intensity and presence score. A list of included action units along with their meaning and related emotions are presented in Table 4.3

Table 4.3: *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 & intensity

4.3.1 Model fitting and ANOVA

The three models in Equation 3.3, 3.4 and 3.5 has been fitted and ANOVA test is performed to choose the best model. In Table 4.4, the deviance of model 1 and 2 are compared and the p-value rejects the null hypothesis that model 1 and model 2 are the same. The comparison of model 2 and model 3 is presented in Table 4.5. We take a conservative approach and not rejecting the null hypothesis that model 3 is too much different from model 2. Thus, model 2 is chosen as our final model for modelling presence.

Table 4.4: *this is the caption*

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
30320	35850	NA	NA	NA
30259	35256	61	594.4	1.793e-88

Table 4.5: *this is the caption*

Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
30259	35256	NA	NA	NA
30253	35242	6	14.21	0.02742

4.3.2 Residual Diagnostics and post-model analysis

Residual analysis is performed on model 2 to illustrate the fitness of the model. In Figure 4.6, the left panel shows the residual for each judge in a boxplot and one can observe that the residuals are around zero for most of the Justices. On the right panel, the residuals are plotted as dots and there is no pattern shown on the plot. This indicates model 2 has a good fit.

The estimated marginal mean is computed in Table A.2 in the Appendix 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
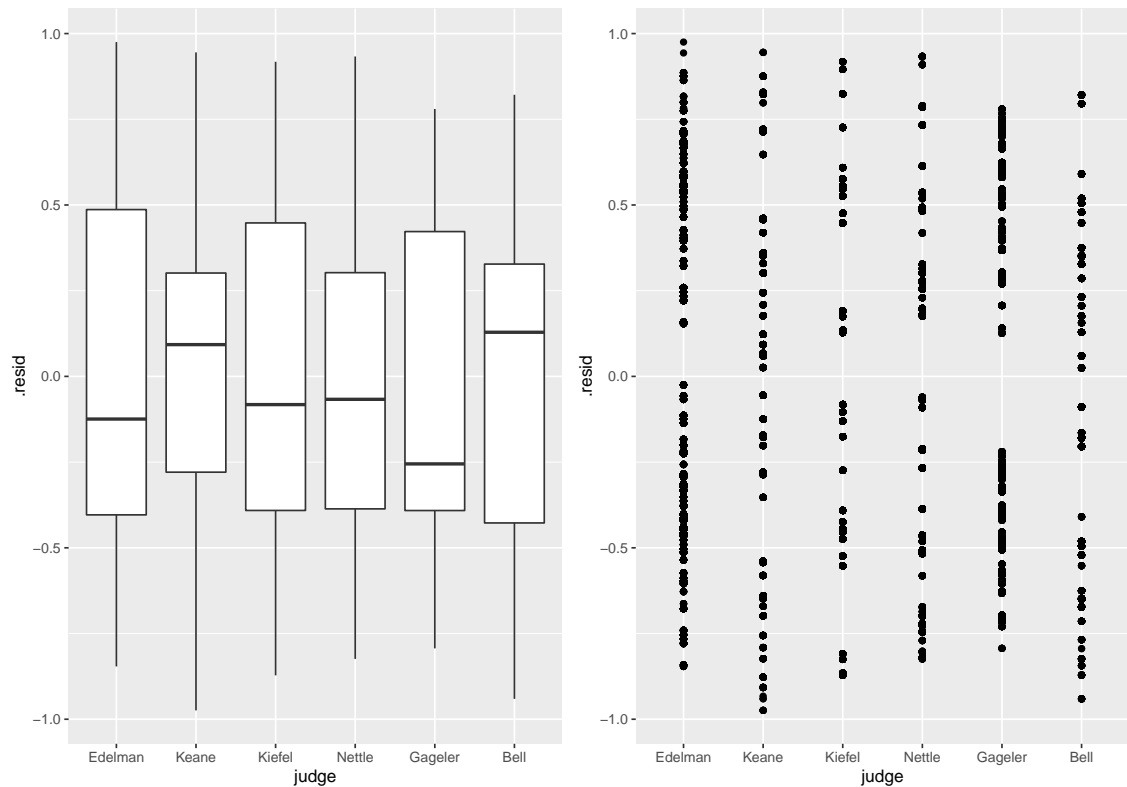
Figure 4.6: *The residual plot for model 2.*

is asymmetric around the estimates because the linear symmetric interval of the mean need to be transferred via the inverse of link function to get the confidence interval for the response. The 95% confidence interval after bonferroni adjustment is plotted in Figure 4.7.

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 4.8.
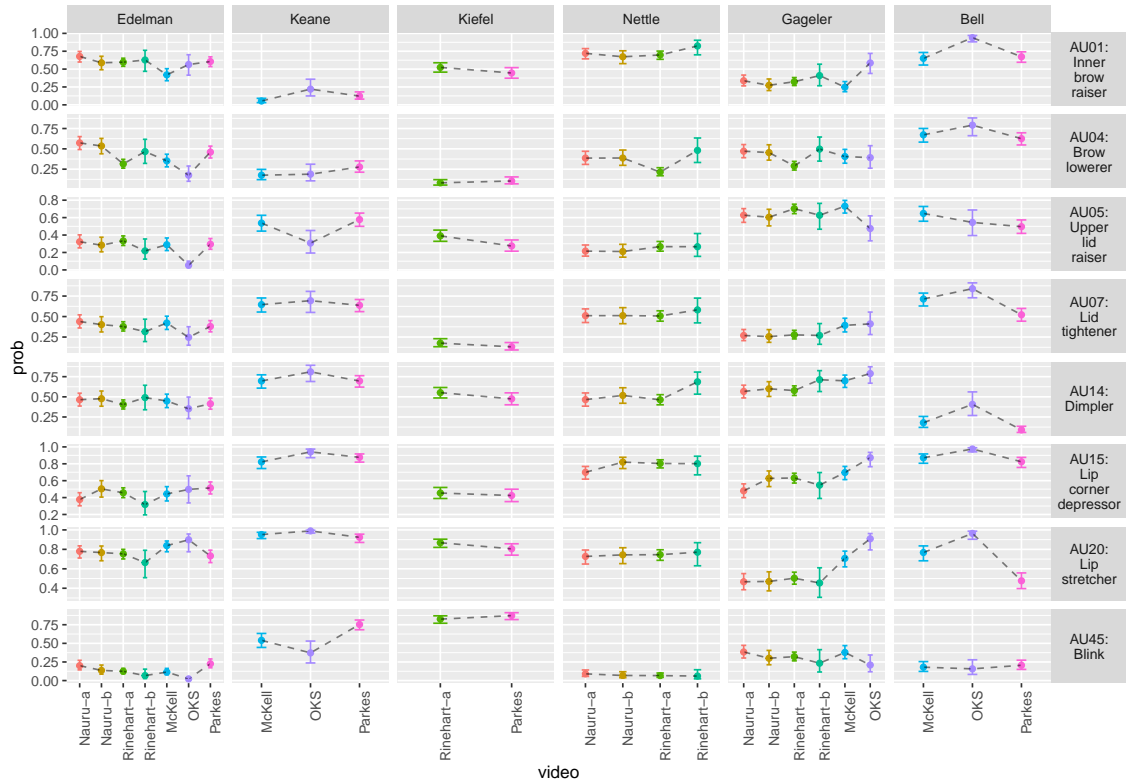
Figure 4.7: *The confidence interval for estimated mearginal mean in model 2*
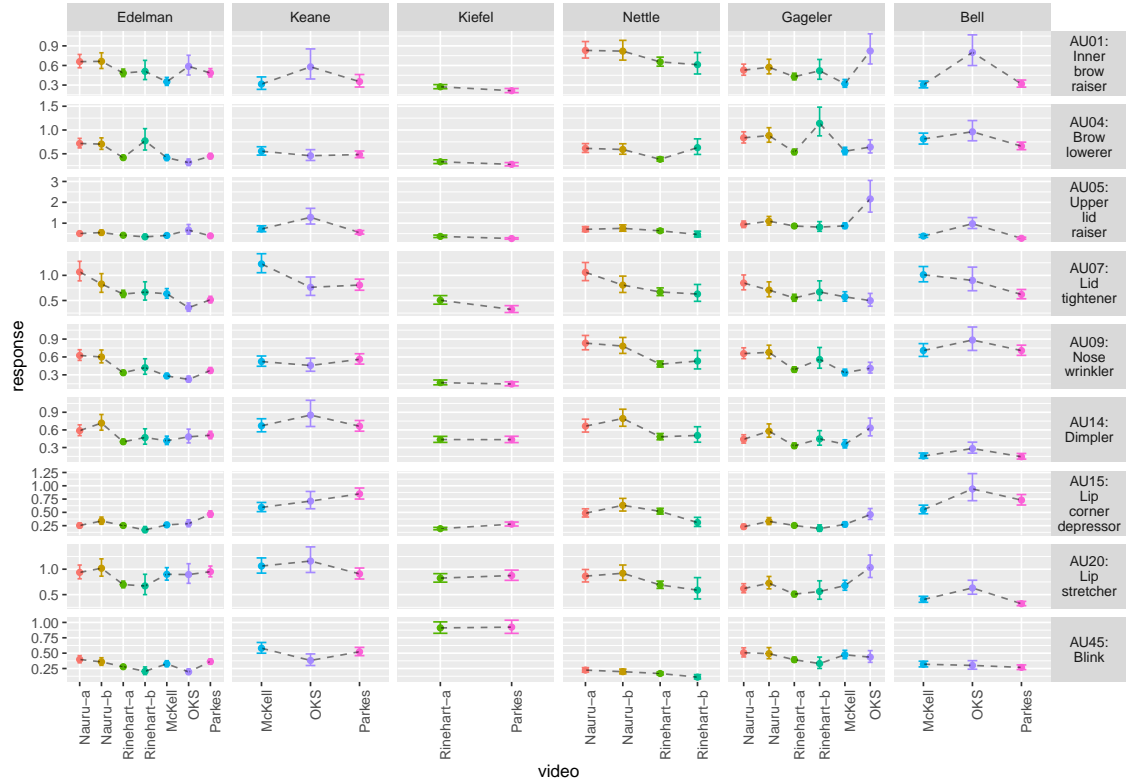
Figure 4.8: *The confidence interval for estimated mearginal mean in model 2*

4.4 Discussion

4.4.1 The expression of the justices by video

Many information can be observed from Figure 4.7 for presence and Figure 4.8. The discussion of the two plots will be focusing on answering the question: **For the same judge, whether the mean presence and mean intensity of the action units are the same or different for different videos**.

In general, we can see that facial expression of the justices are impartial as most of the 95% confidence intervals for the same judge and action unit overlaps in most of the videos in both plots. While at the same time, there are some time when in a particular video, a judge expressed significantly more or less of an expression and I will break it down to discuss them in more detail.

We can observe that Judge Edelman and Keane behave relatively consistent throughout all the videos since most of the intervals overlap with each other after the bonferroni adjustment. However, Edelman seems to express significantly less in action unit 5 in case OKS. In terms of intensity, Edelman has significantly stronger expression in case Nauru-a, Nauru-b and Rinehart-a in action unit 4 and 9, but more relaxed expression in action unit 7 for case OKS.

To be able to draw some conclusion about the facial behaviour of the Justices, it is necessary to have an understanding about the characteristics of the cases. Nauru-a and Nauru-b discuss the refugee status of the appellants, which can be though of political cases. McKell and OKS are more criminal cases, relates to drug and sexual misconduct. Parkes is a civil case about negligence while Rinehart-a and Rinehart-b are business cases. Based the nature of the 7 cases, we can see Edelman is more likely to express stronger emotion to the political and commercial cases and express less and softer at criminal cases.

For Keane, he appears significantly less action unit 5 in case OKS as Edelman in Figure 4.7 but Figure 4.8 suggest he has more intense of expression in the same action unit 5. Keane has also more intense expression of action unit 7 in case McKell. This would show that in general Keane is more responsive to the criminal cases than cases of other categories.

Kiefel and Nettle are relatively consistent in their expressions. The presence plot shows that Nettle has relatively higher mean in case Rinehart-b, but it is not significant when comparing the 95% confidence interval.

Gageler shows a consistent responsive expression in case OKS. In Figure 4.7, Gageler had significantly more expression of action unit 1, 15 and 20. From the intensity plot in Figure 4.8, action unit 5 and 20 have significantly higher intensity. The common associated emotion of action unit 5, 15 and 20 is anger, sad and fear in the six universal emotions. Thus, we can see that Gageler tends to react negatively and more strongly to criminal cases.

Bell also shows a significantly higher proportion of negative emotions associated with action unit 1 and 15 in case OKS and the intensity of action unit 1, 5, 14 and 20 are also significantly higher in case OKS. These evidence indicate the same emotion pattern as judge Gageler to criminal cases. While at the same time, Bell is less reactive in the presence of action unit 07 and 20 in case Parkes and this shows that she has less negative emotion of anger and fear in Parkes, which is the only civil case in our sample videos.

4.4.2 The expression of the justices by speaker

From the presence and intensity plot by speakers in Figure A.1 and A.2, we can observe that the video-wise difference between judge are still persist 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 judges.

This result would be a validation that on the high court level, the judges are behaving impartial to different speaking parties.

4.5 Summary

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

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 apporach 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 reaserach by providing a new, statistical methodology to understand the emotion of the Justices. The facial information gained from this reserach could also be incorporated with other judicial information to predict the high court case outcome in Australia.

5.1 Limitation

I will now briefly discuss some of the limitation of this work. 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.

5.2 Future work

In the future, more work could be done to extract facial expressions of the Justices from videos using OpenFace. This could enable the researchers to capture more precise expression 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.

5.3 Acknowledgement

The analysis is conducted using R [**??**, and the following packages: []

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 Model estimation result

Table A.2: *model result 2 for mean presence*

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	0.417
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	0.362
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	0.698	0.0215	0.637	0.752
Gageler	Rinehart-a	AU01	0.324	0.0208	0.270	0.382
Edelman	Rinehart-b	AU01	0.628	0.0561	0.469	0.763
Nettle	Rinehart-b	AU01	0.824	0.0376	0.700	0.904

Table A.2: *model result 2 for mean presence*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Gageler	Rinehart-b	AU01	0.409	0.0577	0.267	0.569
Edelman	McKell	AU01	0.419	0.0316	0.337	0.506
Keane	McKell	AU01	0.054	0.0107	0.031	0.091
Gageler	McKell	AU01	0.248	0.0264	0.184	0.326
Bell	McKell	AU01	0.650	0.0330	0.557	0.733
Edelman	OKS	AU01	0.564	0.0546	0.415	0.701
Keane	OKS	AU01	0.220	0.0436	0.124	0.358
Gageler	OKS	AU01	0.586	0.0535	0.439	0.719
Bell	OKS	AU01	0.940	0.0156	0.882	0.970
Edelman	Parkes	AU01	0.606	0.0255	0.535	0.672
Keane	Parkes	AU01	0.122	0.0185	0.080	0.181
Kiefel	Parkes	AU01	0.446	0.0277	0.373	0.521
Bell	Parkes	AU01	0.673	0.0273	0.596	0.742
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	0.553
Edelman	Nauru-b	AU04	0.535	0.0356	0.439	0.628
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	0.314	0.0202	0.263	0.371
Kiefel	Rinehart-a	AU04	0.082	0.0122	0.055	0.121
Nettle	Rinehart-a	AU04	0.215	0.0184	0.169	0.268
Gageler	Rinehart-a	AU04	0.290	0.0199	0.239	0.346
Edelman	Rinehart-b	AU04	0.466	0.0567	0.321	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
Edelman	McKell	AU04	0.352	0.0291	0.279	0.434

Table A.2: *model result 2 for mean presence*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Keane	McKell	AU04	0.176	0.0233	0.121	0.248
Gageler	McKell	AU04	0.406	0.0319	0.324	0.493
Bell	McKell	AU04	0.672	0.0315	0.583	0.751
Edelman	OKS	AU04	0.176	0.0346	0.101	0.289
Keane	OKS	AU04	0.189	0.0378	0.107	0.311
Gageler	OKS	AU04	0.392	0.0527	0.262	0.539
Bell	OKS	AU04	0.791	0.0406	0.662	0.880
Edelman	Parkes	AU04	0.460	0.0273	0.388	0.533
Keane	Parkes	AU04	0.277	0.0259	0.213	0.352
Kiefel	Parkes	AU04	0.105	0.0156	0.069	0.155
Bell	Parkes	AU04	0.626	0.0280	0.548	0.697
Edelman	Nauru-a	AU05	0.323	0.0280	0.252	0.402
Nettle	Nauru-a	AU05	0.215	0.0237	0.158	0.286
Gageler	Nauru-a	AU05	0.628	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	0.146	0.294
Gageler	Nauru-b	AU05	0.604	0.0358	0.505	0.696
Edelman	Rinehart-a	AU05	0.333	0.0204	0.280	0.390
Kiefel	Rinehart-a	AU05	0.391	0.0240	0.328	0.457
Nettle	Rinehart-a	AU05	0.267	0.0207	0.215	0.326
Gageler	Rinehart-a	AU05	0.702	0.0202	0.645	0.754
Edelman	Rinehart-b	AU05	0.218	0.0430	0.124	0.355
Nettle	Rinehart-b	AU05	0.266	0.0493	0.155	0.417
Gageler	Rinehart-b	AU05	0.627	0.0567	0.467	0.763
Edelman	McKell	AU05	0.289	0.0269	0.222	0.366
Keane	McKell	AU05	0.537	0.0341	0.445	0.626
Gageler	McKell	AU05	0.731	0.0271	0.652	0.797

Table A.2: *model result 2 for mean presence*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Bell	McKell	AU05	0.648	0.0317	0.559	0.728
Edelman	OKS	AU05	0.053	0.0129	0.027	0.101
Keane	OKS	AU05	0.308	0.0490	0.194	0.452
Gageler	OKS	AU05	0.475	0.0548	0.334	0.620
Bell	OKS	AU05	0.545	0.0560	0.395	0.687
Edelman	Parkes	AU05	0.295	0.0231	0.237	0.360
Keane	Parkes	AU05	0.578	0.0284	0.500	0.652
Kiefel	Parkes	AU05	0.275	0.0240	0.215	0.344
Bell	Parkes	AU05	0.496	0.0289	0.419	0.573
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	0.175	0.0182	0.132	0.230
Nettle	Rinehart-a	AU07	0.507	0.0237	0.443	0.570
Gageler	Rinehart-a	AU07	0.277	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	0.422	0.0304	0.343	0.505
Keane	McKell	AU07	0.646	0.0321	0.555	0.727
Gageler	McKell	AU07	0.393	0.0314	0.312	0.480
Bell	McKell	AU07	0.714	0.0294	0.629	0.786
Edelman	OKS	AU07	0.246	0.0422	0.150	0.376

Table A.2: *model result 2 for mean presence*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Keane	OKS	AU07	0.694	0.0485	0.550	0.807
Gageler	OKS	AU07	0.411	0.0522	0.281	0.554
Bell	OKS	AU07	0.840	0.0334	0.729	0.911
Edelman	Parkes	AU07	0.380	0.0257	0.313	0.451
Keane	Parkes	AU07	0.637	0.0276	0.560	0.708
Kiefel	Parkes	AU07	0.131	0.0167	0.092	0.183
Bell	Parkes	AU07	0.522	0.0291	0.444	0.599
Edelman	Nauru-a	AU14	0.465	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	0.462
Kiefel	Rinehart-a	AU14	0.552	0.0243	0.486	0.616
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	0.697	0.0310	0.608	0.773
Gageler	McKell	AU14	0.699	0.0285	0.617	0.770
Bell	McKell	AU14	0.179	0.0251	0.121	0.257
Edelman	OKS	AU14	0.352	0.0507	0.230	0.497
Keane	OKS	AU14	0.810	0.0372	0.690	0.891
Gageler	OKS	AU14	0.789	0.0379	0.669	0.873

Table A.2: *model result 2 for mean presence*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Bell	OKS	AU14	0.406	0.0560	0.268	0.561
Edelman	Parkes	AU14	0.414	0.0257	0.347	0.484
Keane	Parkes	AU14	0.696	0.0264	0.621	0.762
Kiefel	Parkes	AU14	0.475	0.0278	0.401	0.549
Bell	Parkes	AU14	0.090	0.0139	0.058	0.135
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	0.516
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	0.547	0.0588	0.390	0.696
Edelman	McKell	AU15	0.443	0.0318	0.360	0.530
Keane	McKell	AU15	0.823	0.0253	0.744	0.881
Gageler	McKell	AU15	0.697	0.0291	0.613	0.769
Bell	McKell	AU15	0.871	0.0203	0.806	0.917
Edelman	OKS	AU15	0.497	0.0618	0.337	0.657
Keane	OKS	AU15	0.941	0.0175	0.872	0.974
Gageler	OKS	AU15	0.873	0.0306	0.766	0.935
Bell	OKS	AU15	0.975	0.0080	0.942	0.990
Edelman	Parkes	AU15	0.514	0.0266	0.443	0.585

Table A.2: *model result 2 for mean presence*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Keane	Parkes	AU15	0.876	0.0178	0.820	0.917
Kiefel	Parkes	AU15	0.424	0.0275	0.353	0.499
Bell	Parkes	AU15	0.824	0.0220	0.757	0.876
Edelman	Nauru-a	AU20	0.780	0.0233	0.711	0.836
Nettle	Nauru-a	AU20	0.727	0.0269	0.649	0.793
Gageler	Nauru-a	AU20	0.467	0.0311	0.385	0.551
Edelman	Nauru-b	AU20	0.767	0.0282	0.683	0.834
Nettle	Nauru-b	AU20	0.743	0.0305	0.653	0.817
Gageler	Nauru-b	AU20	0.470	0.0369	0.373	0.569
Edelman	Rinehart-a	AU20	0.754	0.0183	0.701	0.800
Kiefel	Rinehart-a	AU20	0.868	0.0156	0.820	0.905
Nettle	Rinehart-a	AU20	0.745	0.0202	0.687	0.796
Gageler	Rinehart-a	AU20	0.503	0.0230	0.442	0.565
Edelman	Rinehart-b	AU20	0.664	0.0536	0.509	0.790
Nettle	Rinehart-b	AU20	0.771	0.0444	0.631	0.869
Gageler	Rinehart-b	AU20	0.454	0.0583	0.306	0.611
Edelman	McKell	AU20	0.838	0.0205	0.775	0.886
Keane	McKell	AU20	0.952	0.0113	0.911	0.975
Gageler	McKell	AU20	0.707	0.0302	0.620	0.782
Bell	McKell	AU20	0.768	0.0285	0.683	0.835
Edelman	OKS	AU20	0.898	0.0320	0.775	0.958
Keane	OKS	AU20	0.990	0.0045	0.967	0.997
Gageler	OKS	AU20	0.909	0.0292	0.794	0.963
Bell	OKS	AU20	0.964	0.0135	0.904	0.987
Edelman	Parkes	AU20	0.732	0.0238	0.664	0.791
Keane	Parkes	AU20	0.924	0.0153	0.871	0.956
Kiefel	Parkes	AU20	0.806	0.0217	0.740	0.857

Table A.2: *model result 2 for mean presence*

judge	video	AU	prob	SE	asymp.LCL	asymp.UCL
Bell	Parkes	AU20	0.477	0.0302	0.397	0.558
Edelman	Nauru-a	AU45	0.200	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	0.473
Edelman	Nauru-b	AU45	0.136	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	0.869
Nettle	Rinehart-a	AU45	0.067	0.0106	0.043	0.102
Gageler	Rinehart-a	AU45	0.321	0.0220	0.265	0.383
Edelman	Rinehart-b	AU45	0.067	0.0215	0.028	0.154
Nettle	Rinehart-b	AU45	0.061	0.0206	0.024	0.146
Gageler	Rinehart-b	AU45	0.233	0.0564	0.115	0.415
Edelman	McKell	AU45	0.114	0.0157	0.078	0.164
Keane	McKell	AU45	0.540	0.0353	0.445	0.633
Gageler	McKell	AU45	0.378	0.0330	0.294	0.470
Bell	McKell	AU45	0.180	0.0244	0.123	0.255
Edelman	OKS	AU45	0.023	0.0065	0.011	0.048
Keane	OKS	AU45	0.372	0.0562	0.237	0.531
Gageler	OKS	AU45	0.209	0.0420	0.118	0.344
Bell	OKS	AU45	0.158	0.0359	0.083	0.279
Edelman	Parkes	AU45	0.226	0.0221	0.172	0.291
Keane	Parkes	AU45	0.754	0.0243	0.682	0.813
Kiefel	Parkes	AU45	0.872	0.0170	0.819	0.912
Bell	Parkes	AU45	0.205	0.0230	0.150	0.274
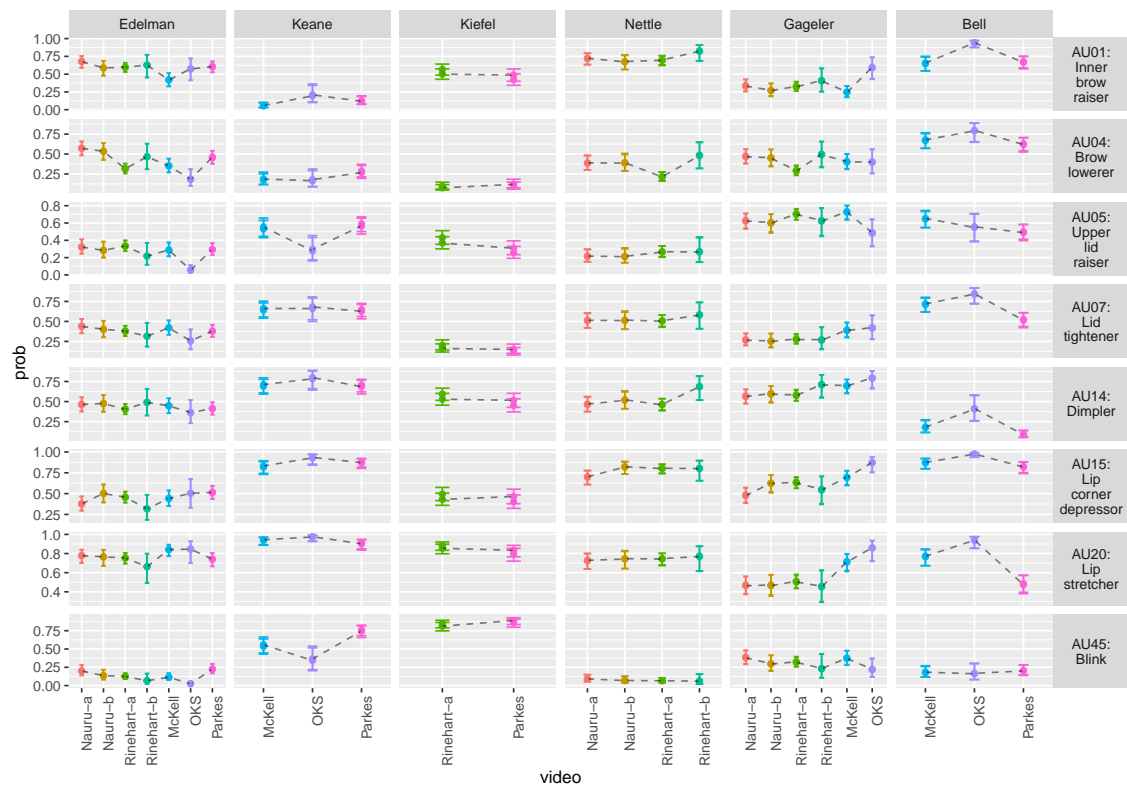
Figure A.1: *The confidence interval for estimated mearginal mean in model 3*
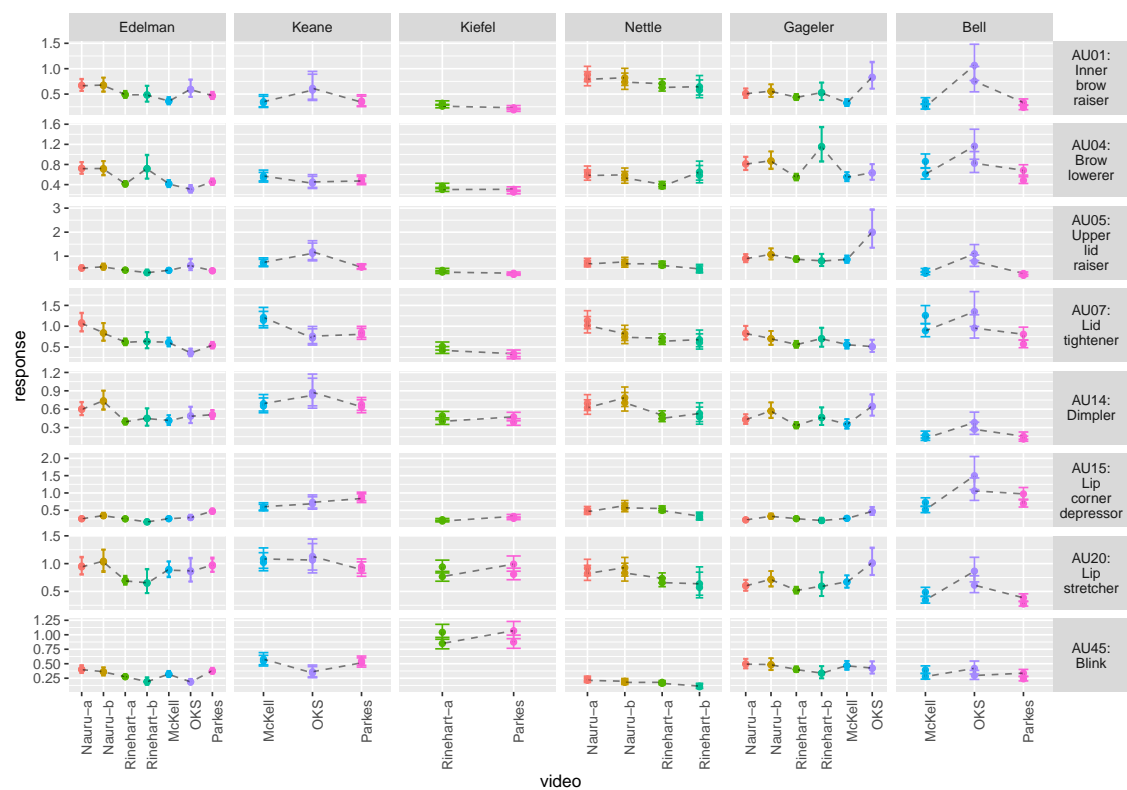
Figure A.2: *The confidence interval for estimated mearginal mean in model 3*

Bibliography

Aliotta, JM (1987-1988). Combining Judges' Attributes and Case Characteristics: An Alternative Approach to Explaining Supreme Court Decisionmaking. *Judicature* **71**, 277.

Australia, HC of (2019). *Recent AV recordings*. Accessed: 2019-05-03. `http://www.hcourt.gov.au/cases/recent-av-recordings`.

Baltrusaitis, T, A Zadeh, YC Lim, and LP 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.

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

Black, RC, SA Treul, TR Johnson, and J Goldman (2011). Emotions, oral arguments, and Supreme Court decision making. *The Journal of Politics* **73**(2), 572–581.

Chen, DL, M Kumar, V Motwani, and P Yeres (2018). *Is Justice Really Blind? And Is It Also Deaf*. Tech. rep. Technical report.

Chen, D, Y Halberstam, and C Alan (2016). Perceived masculinity predicts us supreme court outcomes. *PloS one* **11**(10), e0164324.

Chen, D, Y Halberstam, A Yu, et al. (2017). Covering: Mutable characteristics and perceptions of voice in the US Supreme Court. *Review of Economic Studies invited to resubmit, TSE Working Paper* (16-680).

Chief Justices of Australia, TC of and N Zealand (2017). *Guide to Judicial Conduct*. 3rd. Melbourne: Austral-asian Institute of Judicial Administration.

Cohn, JF, TS Kruez, I Matthews, Y Yang, MH Nguyen, MT Padilla, F Zhou, and F De la Torre (2009). Detecting depression from facial actions and vocal prosody. In: *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp.1–7.

Cragg, JG (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-dl*. `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, BJ, RD Enos, and M Sen (2019). Emotional arousal predicts voting on the US supreme court. *Political Analysis* **27**(2), 237–243.

Duan, N, WG Manning, CN Morris, and JP Newhouse (1984). Choosing between the sample-selection model and the multi-part model. *Journal of Business & Economic Statistics* **2**(3), 283–289.

Ekman, P and WV Friesen (1976). Measuring facial movement. *Environmental psychology and nonverbal behavior* **1**(1), 56–75.

Ekman, P, WV Friesen, and JC Hager (2002). Facial action coding system: The manual on CD ROM. *A Human Face, Salt Lake City*, 77–254.

Ekman, P, M O'Sullivan, WV Friesen, and KR Scherer (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of nonverbal behavior* **15**(2), 125–135.

Ekman, P and WV riesen (1978). *Facial action coding system*. Palo Alto: CA: Consulting Psychologists Press.

Epstein, L, WM Landes, and RA 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, JJ (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, YC, R Amine, and M Sergey (2019). *youtube-dl*. `https://github.com/ytdl-org/youtube-dl/`.

Huang, CL and YM 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*. CHI '18. Montreal QC, Canada: ACM, pp.277:1–277:12. http://doi.acm.org/10.1145/3173574.3173851.

Johnson, TR, RC Black, J Goldman, and SA Treul (2009). Inquiring minds want to know: Do justices tip their hands with questions at oral argument in the US supreme court. *Washington University Journal of Law and Policy* **29**, 241.

Kapoor, A, Y Qi, and RW Picard (2003). Fully automatic upper facial action recognition. In: *IEEE International SOI Conference. Proceedings (Cat. No.03CH37443)*, pp.195–202.

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.

Kobayashi, H and F Hara (1992). Recognition of Six basic facial expression and their strength by neural network. In: *Proceedings IEEE International Workshop on Robot and Human Communication*, pp.381–386.

Koppen, PJ van and JT Kate (1984). Individual Differences in Judicial Behavior: Personal Characteristics and Private Law Decision-Making. *Law and Society Review* **18**(2), 225–247.

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, CT and MB Perry Elissa 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.

Leung, SF and S Yu (1996). On the choice between sample selection and two-part models. *Journal of econometrics* **72**(1-2), 197–229.

Lien, JJJ, T Kanade, JF Cohn, and CC Li (2000). Detection, tracking, and classification of action units in facial expression. *Robotics and Autonomous Systems* **31**(3), 131–146.

Liu, L, RL Strawderman, ME Cowen, and YCT Shih (2010). A flexible two-part random effects model for correlated medical costs. *Journal of Health Economics* **29**(1), 110–123.

Lucey, P, JF Cohn, T Kanade, J Saragih, Z Ambadar, and I Matthews (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp.94–101.

Manning, WG, N Duan, and WH Rogers (1987). Monte Carlo evidence on the choice between sample selection and two-part models. *Journal of econometrics* **35**(1), 59–82.

Manning, WG, CN Morris, JP Newhouse, LL Orr, N Duan, EB Keeler, A Leibowitz, KH Marquis, MS Marquis, and CE Phelps (1981). A two-part model of the demand for medical care: preliminary results from the health insurance study. *Health, economics, and health economics*, 103–123.

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.

Nasir, M, A Jati, PG Shivakumar, S Nallan Chakravarthula, and P Georgiou (2016a). Multi-modal and Multiresolution Depression Detection from Speech and Facial Landmark Features. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. AVEC '16. Amsterdam, The Netherlands: ACM, pp.43–50. http://doi.acm.org/10.1145/2988257.2988261.

Nasir, M, A Jati, PG Shivakumar, S Nallan Chakravarthula, and P Georgiou (2016b). Multimodal and multiresolution depression detection from speech and facial landmark features. In: *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, pp.43–50.

Neelon, B and AJ O'Malley (2019). "Two-Part Models for Zero-Modified Count and Semicontinuous Data". In: *Health Services Evaluation*. New York, NY: Springer US, pp. 695–716. https://doi.org/10.1007/978-1-4939-8715-3_39.

Pan, X and AFdC Hamilton (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology* **109**(3), 395–417. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bjop.12290.

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, JN, SA 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, SL (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 CL 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.

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.

Welch, S, M Combs, and J Gruhl (1988). Do Black Judges Make a Difference? *American Journal of Political Science* **32**(1), 126–136.

Yang, L, D Jiang, L He, E Pei, MC 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.