Factors influencing affective acuity

A report submitted for the course
COMP4560 Advanced Computing Project

Author: Xue Ling Teh (u6462117)
12 June 2020

Supervisors:
Prof Tom Gedeon
Dr Sabrina Caldwell
Jessica Rahman
Declaration

To the best of the author’s knowledge, this report does not contain any material previously published or written by another person, except where due reference is made in the text.

Xue Ling Teh

June 2020
Acknowledgements

I am enormously grateful to my supervisors Professor Tom Gedeon, Dr Sabrina Caldwell, and Jessica Rahman, for their continuous support and guidance throughout the stages of this project. I wish to express my gratitude to the rest of the Human Centred Computing (HCC) team, including Ruimin Chu, Yang Liu, Zi Jin, and Amelia Poetter for their contribution in the experiments’ success. Additionally, I would like to pay my special regards to Jessica and Zi for sharing their code, knowledge and expertise from previous related areas of research, and Ruimin for sharing her discoveries on the participants who watched the Perception of Truth and Uncertain Statements (POTUS) videos in our shared experiments.

I would also like to thank my parents, Kok Leong Teh and Eliza Lim, and my friends for the constant encouragement and support throughout my studies and Honours project.

Special thanks to the course convenor Professor Weifa Liang for general tips on managing the Honours project and ANU Academic Skills for advice on thesis report writing.

To all the participants who have been a part of the experiment – thank you for your time and effort in helping me complete my research.

This research has been approved by the ANU Human Research Ethics Committee (Protocol 2018/489). I wish to thank the ANU Research School of Computer Science for the research opportunity and in providing the equipment and experiment environment to conduct this study as part of my Honours project.
Abstract

Recent advances in the Human Centred Computing (HCC) field have inspired research studies within the subfield of Affective Computing (AC) on affective, or emotional reasoning, perception and judgment. Existing research has uncovered the benefits of exercising and drawbacks of caffeine intake on human creativity and cognitive tasks. However, the effects of exercise and caffeine on affective perception and reasoning has not been extensively researched. This study investigates two potential factors or prior conditions which influence human affective perception and reasoning. The two major conditions of interest are exercising by walking and caffeine intake from drinking mostly coffee. An additional control condition is observed, in which participants were provided with a magazine to read. The two-part experiment includes the completion of one of the activity conditions, followed by watching short clips of emotion-based videos and answering related questions. The participant’s emotional reasoning and judgment of the emotions portrayed in the videos, whether consciously or unconsciously, indirectly measures whether the participant believes in the statements made by the people in the video. Throughout the experiment, participants’ physiological signals are collected. The main physiological signal which was analysed was the Electrodermal Activity (EDA). Each participant’s EDA signal is compared against the ground truth information of the videos that they watched. Results are obtained by training and testing a neural network classification model with the selected statistical summary of EDA features. According to the mean classification accuracy results, exercising has a positive effect on the ability to emotionally reason, whereas having caffeine generally led to poorer performance in comparison. A somewhat surprising discovery was that male participants performed better than the female cohort, which contrasts the traditional belief that females are better at handling emotion-related tasks. These findings provide supporting evidence to the reliability of future research results, depending on the prior conditions and factors in an emotion-related experiment based on a participant’s recent physical activity, gender, ethnicity background. In the future, more in-depth research on affective reasoning can be conducted by analysing other types of physiological signals, and by investigating different combinations of influencing factors.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance test for statistical significance of results</td>
</tr>
<tr>
<td>AC</td>
<td>Affective Computing</td>
</tr>
<tr>
<td>ANU</td>
<td>Australian National University</td>
</tr>
<tr>
<td>BVP</td>
<td>Blood Volume Pulse</td>
</tr>
<tr>
<td>COVID-19</td>
<td>2019 Novel Coronavirus Disease, formerly known as 2019-nCoV</td>
</tr>
<tr>
<td>CSIT</td>
<td>ANU Computer Science and Information Technology building</td>
</tr>
<tr>
<td>DNF</td>
<td>DNF Dunbar Physics Lecture Theatre located at the ANU</td>
</tr>
<tr>
<td>EDA</td>
<td>Electrodermal Activity</td>
</tr>
<tr>
<td>ERQ</td>
<td>Emotion Regulation Questionnaire</td>
</tr>
<tr>
<td>GAU</td>
<td>Guilford’s Alternate Uses</td>
</tr>
<tr>
<td>GSR</td>
<td>Galvanic Skin Response</td>
</tr>
<tr>
<td>HCC</td>
<td>Human Centred Computing</td>
</tr>
<tr>
<td>HCI</td>
<td>Human Computer Interaction</td>
</tr>
<tr>
<td>HRV</td>
<td>Heart Rate Variability</td>
</tr>
<tr>
<td>ID</td>
<td>Identifier</td>
</tr>
<tr>
<td>POTUS</td>
<td>Perception of Truth and Uncertain Statements dataset</td>
</tr>
<tr>
<td>SONA</td>
<td>Online Research Participation Management</td>
</tr>
<tr>
<td>ST</td>
<td>Skin Temperature</td>
</tr>
<tr>
<td>WATTLE</td>
<td>Web Access To Teaching &amp; Learning Environments</td>
</tr>
</tbody>
</table>
# Table of Contents

Declaration ....................................................................................................................... iii
Acknowledgements ........................................................................................................ v
Abstract ........................................................................................................................ vii
List of Abbreviations ........................................................................................................ ix
Table of Contents ........................................................................................................... xi
List of Tables ..................................................................................................................... xiii
List of Figures ................................................................................................................... xv
Chapter 1: Introduction ...................................................................................................... 1
Chapter 2: Background Literature .................................................................................... 3
  2.1. Current state-of-the-art .......................................................................................... 3
  2.2. Project Scope and Definition .............................................................................. 4
Chapter 3: Experiment Design and Analysis .................................................................. 6
  3.1. Emotion and Empathy Experiment Design ......................................................... 6
  3.2. Conducting the Experiment .................................................................................. 8
  3.3. Proposed Analysis ................................................................................................. 10
Chapter 4: Classification Results and Discussion ............................................................ 12
  4.1. Main Experiment (POTUS) .................................................................................. 12
  4.2. Pilot study (EmotiW) ............................................................................................ 17
Chapter 5: Conclusion ...................................................................................................... 18
  5.1 Future work ........................................................................................................... 18
References ....................................................................................................................... 20
Appendix 1 ....................................................................................................................... 23
Appendix 2 ....................................................................................................................... 24
Appendix 3 ....................................................................................................................... 25
Appendix 4 ....................................................................................................................... 32
List of Tables

Table 1: Demographic Breakdown of POTUS Participants.......................................................... 9
Table 2: Segmentation of POTUS Participants by Categories .................................................. 11
Table 3: Summary Statistics of POTUS Classification Percentage Accuracy for 20 Iterations Based on Categories .............................................................................................................. 12
Table 4: Summary Statistics of POTUS Classification Percentage Accuracy for 20 Iterations Based on 13-sized subsets .............................................................................................................. 15
List of Figures

Figure 1: The Empatica E4 wristband (Source: Empatica) ................................................................. 7
Figure 2: Eye-calibration on TheEyeTribe Eye-tracking Device (Source: TheEyeTribe) .............. 7
Figure 3: Demographic of POTUS Participants for 3 Experiment Conditions ......................... 10
Figure 4: Standard Deviation of POTUS Classification Accuracy Based on 8 Categories ........ 13
Figure 5: POTUS Classification Percentage Accuracy for 20 Iterations Based on 8 Categories ................................................................. 14
Figure 6: POTUS Classification Percentage Accuracy for 20 Iterations Based on Statistical Summary ................................................................. 14
Figure 7: POTUS Classification Mean Percentage Accuracy for 20 Iterations Based on Segmented Subsets ................................................................. 16
Chapter 1: Introduction

The Human Centred Computing (HCC) field exists to bridge the existing gaps between various disciplines, including the human sciences, such as psychology, and computer science [8], [10], [17]. One of the main goals of HCC revolves around improving the technological designs of computing systems and appliances for humans. Closely related to HCC is Human Computer Interaction (HCI), which focuses on studying how human users interact with data and the usability of systems. Interdisciplinary research experiments conducted in HCC and HCI often incorporate literature from various fields such as computing, psychology, and biologically inspired computing. These experiments are carried out using wearable and mobile devices to collect affective data in the form of video recordings, eye-gaze and physiological signals [26], [33].

Research in HCC has grown over the last two decades and studies by scholars on Affective Computing (AC) aim to understand the different ways that human emotions have an impact on our reasoning, perception and judgment [3], [22], [30]. In fact, the word ‘affect’ can be considered equivalent to subjective human emotions. The most widely used characterisation of emotions by psychologists is categorising the basic emotions as follows: happiness, sadness, fear, anger, disgust, and surprise [33].

The term “affective computing” was an emerging term towards the end of the 20th century, which was defined by Picard as “computing that relates to, arises from, or influences emotions” [22]. Picard surmised that future computers would be integrated into modern society, assuming roles such as a tutor or learning advisor which are sensitive to human emotional needs. One of the core aspects of AC is affect detection [2], which detects or senses the underlying human emotions of users, and affective perception [32]. Another study by Vesterinen noted the difference between recognising or expressing acted emotions and having or feeling actual emotions [30]. Various research findings in AC enables us to create better decision-making systems which could have emotional intelligence [22], [30].

Despite existing research on affective reasoning and perception [1], [30], [32], [33], the comparison of positive and negative influencing factors on affective perception has not been thoroughly researched. This research project aims to investigate the potential factors that influence human emotional perception and reasoning – specifically on exercise and caffeine as prior conditions. The objective is to find out which of these factors would have an effect, whether directly or indirectly, on people's emotional judgment.
The remainder of this report is divided into four sections (chapters). Chapter 2 describes the related studies in the literature to provide a background context and the inspiration behind this research project. In Chapter 3, details of designing and conducting the experiment and the proposed analysis of data are elaborated. Results from training and testing the classification model on the gathered data are described in Chapter 4, which leads to the derived conclusions, including potential future works in the final chapter (Chapter 5).
Chapter 2: Background Literature

Existing literature supports the notion that physical activity such as exercise, which includes walking, is linked to human creativity and encourages productivity. In contrast, there are different views from studies on the effects of caffeine as stimuli. The gender and ethnicity of participants are also of interest in the analysis of affective perception. The current state-of-the-art section highlights some of the benefits of exercise and the effects of caffeine intake on human cognition and the consideration of gender and ethnicity in which inspired the foundation of my research project.

2.1. Current state-of-the-art

Several studies agreed that walking has a positive effect on creativity. In Oppezzo and Schwartz’s research [16], one of the measures used was Guilford’s alternate uses (GAU) test of creative divergent thinking. They noted that walking resulted in higher GAU scores, promotes creative thought process in real-time and for a short period afterwards. Similarly, the results of Zimmermann’s research [34] indicate that regular physical activity helps in drawing people’s attention to relevant information and disregarding irrelevant information when making product judgments. With physical activity, consumers place a higher consideration in product feasibility over desirability when faced with an evident compromise between the two features. Walking is also recommended in trauma counselling as a good way to cope with grief, which is detailed in Clark’s recent investigation on the effects of walk-and-talk therapy [4] and Megas’ study on walking as a counselling practice for grief and trauma [14]. According to [4], walk-and-talk therapy is defined as therapy sessions which incorporate physical activity such as walking in an outdoor environment. Clark has found that clients deem the alternative walk-and-talk therapy as therapeutic as, or more so than traditional therapy and felt that the former could be a less stigmatising option for individuals who were not keen on traditional, indoor therapy [4].

On the other hand, several studies found that exercising by walking has little to no effect on affective decision-making. Opdal’s experiment in 2015 [15] led to the conclusion that there was no significant effect in mind wandering, which is comparable to daydreaming, and creativity between students who walked or rested. Another study by Stanley, Elrod, Twyman and Zou [29], which explores the influence of walking and mobile shopping simultaneously, seem to share the same sentiment. Whether customers walked or remained seated while mobile shopping did not change the amount of money spent, which indicates that the customers’ belief in brands did not waver. Although these findings seem to suggest that walking may not have
an effect, trust in brandings of the same product and recognising the degree of doubt on faces are very different in nature and likely to utilise different parts of the brain.

Many studies support the advantages of caffeine, such as inducing alertness and enhancing cognition of individuals. [28] states that caffeine improves an individual’s overall mood which leads to better performance in cognitive tasks. According to [7], having coffee in moderation can help people stay awake, giving them a much-needed boost of energy and improve the accuracy of responses in cognitive tasks.

Despite many evidence supporting caffeine alertness, one study in particular by Gu [9] came to a different conclusion when it came to human subjective doubt. Gu [9] discovered that people who had coffee or tea before the experiment tend to believe rather than doubt what was presented to them. His results also show that humans make better judgment unconsciously, which is measured through physiological signals, compared to conscious judgment decisions. Based on his findings, he derived that caffeine was the determining factor which reduced our ability to detect deception.

Studies have shown the consideration of gender and ethnicity as influencing factors in affective perception. In [25], the researcher discusses the importance of uniquely identifying characteristics of computer application users, in terms of gender and ethnicity. For example, a study discovered that females observed a higher score in distinguishing genuine smiles than male participants [12]. Observing gender and ethnicity differences as part of this research would provide valuable insight into how culture shapes our affective perception.

2.2. Project Scope and Definition

Gu’s discovery of caffeine’s negative impact on emotional reasoning piqued my interest and propelled me to investigate further, along with the effects of exercise. The unique focus of my project is to investigate the potential factors that influence human emotional perception and reasoning – specifically on exercise and caffeine as prior conditions. The aim is to find out which of these factors would have an effect, whether directly or indirectly, on people’s emotional judgment. Despite existing research on both positive and negative independent factors, the comparison of positive and negative influencing factors on affective perception has not been thoroughly researched. In this study, we would be able to find out whether there are any short-term effects of exercising and caffeine prior to performing emotional reasoning tasks, and account for the participants’ gender and ethnicity. Therefore, the question this study aims to answer is: What is the extent in which certain prior conditions, such as exercising and
caffeine intake, affect an individual's affective perception, and what can we tell from observing participant gender and ethnicity?
Chapter 3: Experiment Design and Analysis

As discussed in Chapter 2: Background Literature, two potential influencing factors that affect our emotional perception are stimuli such as caffeine intake, and daily activities, such as exercise. Hence, the following 3 main hypotheses are made:

H1: Exercising would have a positive effect and thus participants are predicted to perform better than chance in emotionally recognising or detecting lies from statements.

H2: Caffeine intake would make a person more gullible and believing in statements, which would likely lead to a poorer performance compared to normal conditions.

H3: Under normal circumstances, participants would do better in emotional perception tasks than participants who had drinks containing caffeine, but not as well as participants who had exercised beforehand.

In order to test my hypotheses, I designed an experiment using videos relating to emotions. The novel video dataset which consists of Perception of Truth and Uncertain Statements (POTUS)\(^1\) was chosen for the main experiment, and the Emotion Recognition in the Wild (EmotiW)\(^2\) video dataset was used for the pilot study.

3.1. Emotion and Empathy Experiment Design

The POTUS video dataset used in the main experiment is based on a feature of the American newspaper, The Washington Post, called The Fact Checker\(^3\). The Fact Checker has an ongoing database of short few-second video clips of statements made by USA President Donald Trump across a range of different topics since assuming office, including the fact check ratings which indicate the degree of truth to those statements compiled by a dedicated “truth squad” team. The sixteen topics included in the POTUS videos are biographical record, crime, economy, election, environment, foreign policy, guns, health care, immigration, jobs, miscellaneous, Russia, taxes, terrorism, trade, and Ukraine probe. These videos were order-balanced similarly to the Latin Square design [13], with some additional constraints. For each POTUS topic, there would be at least 1 video (between 1-2 videos) shown to participants in each session, and the order of topics was shuffled by the Latin Square method. Each participant would see one or more videos of the same topic. The number of times each video

\(^1\) https://www.washingtonpost.com/graphics/politics/trump-claims-database/?fbclid=IwAR1wHmsvM1mKcmN21BEXTYuGFlIDmhruOrRdqvv48-8UpYEk_xExdoZh5A
\(^2\) https://sites.google.com/view/emotiw2020
\(^3\) https://www.washingtonpost.com/politics/2019/01/07/about-fact-checker/
was watched during the experiments is balanced so that each video is seen a similar number of times, despite having videos of different lengths.

The variables I tested in my experiment were caffeine intake and exercise by taking a short walk that lasts approximately 8 minutes. The control group was given a magazine to read. Three main types of equipment were used:

- Empatica E4 wristband⁴
- TheEyeTribe eye-tracking device⁵
- Logitech C922 Pro Stream Webcam⁶

![Empatica E4 wristband](https://www.empatica.com/research/e4/)

**Figure 1:** The Empatica E4 wristband (Source: Empatica)

![Eye-calibration on TheEyeTribe Eye-tracking Device](https://theeyetribe.com/theeyetribe.com/about/index.html)

**Figure 2:** Eye-calibration on TheEyeTribe Eye-tracking Device (Source: TheEyeTribe)

⁴ [https://www.empatica.com/research/e4/](https://www.empatica.com/research/e4/)
⁵ [https://theeyetribe.com/theeyetribe.com/about/index.html](https://theeyetribe.com/theeyetribe.com/about/index.html)
The E4 wristband collects the participant’s physiological signals, including the Galvanic Skin Response (GSR) such as Electrodermal Activity (EDA), Heart Rate Variability (HRV), Blood Volume Pulse (BVP) and Skin Temperature (ST). EDA is sampled at 4Hz per second. ST is also sampled at 4Hz. BVP and heart rate are recorded at 64Hz per second. The EyeTribe is an eye-tracking pupillometry device which captures the participant’s pupillary response, with the sample frequency rate set to 60Hz per second. The Logitech webcam, which was connected to a separate laptop, recorded their facial expressions and reactions throughout the experiment. Questionnaire responses from the participant were also collected. An example is the Emotion Regulation Questionnaire (ERQ), which was shown right before the end of the experiment. Participants were recruited from two ANU computing courses via the Web Access To Teaching and Learning Environments (WATTLE) system, ANU’s e-learning Moodle platform, and the SONA system which is an Online Research Participation Management System managed by the ANU Research School of Psychology. Further details on the experiment such as the participant information sheet, consent and questionnaire forms can be found in Appendix 3.

3.2. Conducting the Experiment

The experiment took place in two rooms, N236 and N239, Level 2, Computer Science and Information Technology (CSIT) Building 108 on the ANU Campus. After the participants’ arrival, they are given a short briefing of the study by the researcher(s), then provided with the participant information sheet (Appendix 3.3) and consent form to sign (Appendix 3.1). Before beginning the experiment, they were asked a question of whether they had coffee or tea within the last hour or earlier during the day before coming to the experiment. The researcher also asked the participant for their non-dominant hand and helps the participant in wearing the E4 wristband.

The experiment is divided into 2 main tasks, the total duration of which is approximately 30 minutes. Prior to starting the activity, the button on the participant’s E4 wristband is pressed or ‘tagged’ in order to record the timestamp. For the first 10-15 minutes, participants were guided in doing one of the activities:

1. Either drink coffee or tea (most of them being coffee),
2. Go for a walk, or
3. Read a magazine.
The medium cup coffee provided was from the coffee machine located in the tea room at Level 2 of the CSIT building. The 2 participants who drank tea instead of coffee were given English Twinings tea. The exercising participants were accompanied by the researcher(s), following a predefined walking route which is an approximate 8-min return trip from the CSIT Experiment room to the DNF Dunbar Physics Theatre and back. The magazine provided consists of topics related to science and technology. After the activity has been completed, they continued with the second part of the experiment for the rest of the duration. In the second part, participants were asked to watch a series of short video clips and answered questions related to the portrayed emotions. Participants were seated comfortably in front of a laptop. During the setup, eye calibration was performed for the TheEyeTribe eye-tracking device to track their eye-gaze and pupil dilation at 60Hz, as shown in Figure 2 and previously discussed in the experiment design. The video recording is then started as the participants began the second part of the experiment. Starting with a pre-experiment questionnaire which contained demographic-related questions.

There was a total of 44 participants aged 17-40, all of whom are university students at the ANU, who took part in this experiment. Under the recent circumstances with COVID-19 campus closure and shift to remote online learning mode, I did my best to have an even distribution of demographic data in terms of gender and ethnicity. Out of the total of 44 participants, 41 belong to the main POTUS experiment, and the remaining 3 belong to the pilot study, EmotiW. As shown in Table 1, there were a total of 20 male and 21 female participants for the main POTUS experiment, with a further breakdown of the ethnicity groups (Chinese and Other) within the gender groups. Overall, the initial total for the ethnicity categories were 25 Chinese and 16 Other. For the 3 activities, there were a total of 14 Caffeine, 16 Exercise, and 11 Magazine. A visualisation of the demographics is displayed in Figure 3.

**Table 1**: Demographic Breakdown of POTUS Participants

<table>
<thead>
<tr>
<th>Male (20)</th>
<th>Female (21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese (14)</td>
<td>Other (6)</td>
</tr>
<tr>
<td>C</td>
<td>E</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

C = Caffeine, E = Exercise, M = Magazine
The participant's judgment of the emotions portrayed in the videos, whether consciously or unconsciously, indirectly measures whether the participant believes Trump is telling the truth or lying. The participant's physiological signals give insight to the participant's unconscious judgment and the questionnaire responses are a measure of their active and conscious reasoning. Upon completion of the experiment, the next step is to analyse the collected data. The proposed analysis is discussed in the next section.

3.3. Proposed Analysis

The main idea of the analysis is to compare the participants' physiological responses against the ground truth information of the videos. The type of physiological signal that was analysed is EDA, which is one of the most widely researched response systems [5]. The ground truth for the POTUS videos are their respective fact check ratings.

There were three missing E4 data for the main POTUS dataset (3 Female Caucasians, 1 from each activity condition), therefore the remaining usable data were from 38 participants. Table 2 shows the number of POTUS participants belonging to each of the different category segmentations.
Table 2: Segmentation of POTUS Participants by Categories

<table>
<thead>
<tr>
<th>Segmentation by</th>
<th>Category</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Conditions</td>
<td>Caffeine</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Exercise</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Magazine</td>
<td>10</td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>18</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Chinese</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>13</td>
</tr>
</tbody>
</table>

Note: The three missing E4 data are not included in this table.

The raw data is pre-processed by normalization, filtering, and then selecting the most suitable features to train the neural network classification model. An overview of the steps from pre-processing to the neural network classification are explained as follows.

For each of the segmentation, the EDA data is pre-processed by min-max normalization to the 0-1 range [20]. After normalization, any noisy data are filtered using 10th-order median smoothing filter [6], [21]. Based on each participant’s matching EDA and video timestamps, the filtered EDA data, label-encoded video IDs and the respective videos’ fact check ratings are segmented. Based on the 4 different fact check rating values, each participant’s EDA data are sorted. For feature selection and calculation as input to train the neural network, statistical summaries from the sorted data are derived. One of the statistical summary features is welch, which converts time series data to the frequency (Hz) domain [18]. Leave-one-out cross-validation [31] was used in training and testing the neural network classification model.

After feeding the pre-processed data and training the model, we can find out whether each of the 3 different activity conditions has an effect (improvement, deterioration, no difference) on the participants' emotional reasoning. Additionally, we can evaluate if there are any differences in the model prediction accuracy according to the 3 different activity conditions.
Chapter 4: Classification Results and Discussion

This chapter discusses the classification results obtained, as well as the statistical tests, Analysis of variance (ANOVA) test [11], [19] and T-test [19] performed to evaluate the significance of these results. Based on the statistical significance of the results, we can determine which hypotheses are to be accepted.

4.1. Main Experiment (POTUS)

As the average size of the dataset for a segmented category is approximately 13 participants (as shown in Table 2), the mean accuracy of three 13-sized-subsets for the segmentations which had at least 2 more than the average size was obtained. This was the case for the full POTUS dataset (all), gender (male and female) and Chinese segmentations, which is reflected in Table 3. The classification accuracy with the original size of each of these dataset segmentations are shown in Table 3. For a summarized graph of the average accuracy of the 3 subsets of each category, please refer to Figure 7.

Table 3: Summary Statistics of POTUS Classification Percentage Accuracy for 20 Iterations Based on Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>mean</th>
<th>std</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>86.4</td>
<td>3.1</td>
<td>79.0</td>
<td>91.9</td>
</tr>
<tr>
<td>Caffeine</td>
<td>86.5</td>
<td>4.7</td>
<td>81.0</td>
<td>95.2</td>
</tr>
<tr>
<td>Exercise</td>
<td>90.3</td>
<td>4.3</td>
<td>82.0</td>
<td>96.0</td>
</tr>
<tr>
<td>Magazine</td>
<td>87.0</td>
<td>5.3</td>
<td>76.7</td>
<td>96.7</td>
</tr>
<tr>
<td>Male</td>
<td>88.3</td>
<td>4.3</td>
<td>80.9</td>
<td>96.6</td>
</tr>
<tr>
<td>Female</td>
<td>87.2</td>
<td>4.1</td>
<td>78.3</td>
<td>92.2</td>
</tr>
<tr>
<td>Chinese</td>
<td>88.3</td>
<td>4.1</td>
<td>81.7</td>
<td>96.2</td>
</tr>
<tr>
<td>Other</td>
<td>87.1</td>
<td>2.8</td>
<td>82.1</td>
<td>92.3</td>
</tr>
</tbody>
</table>

After feeding in the 38 pre-processed EDA data for all POTUS participants into the neural network for 20 iterations, the resulting mean accuracy is 86.4%. Similarly, I have further segmented the dataset into other categories as follows, with their respective mean accuracies for 20 iterations:

- 3 conditions
  - Caffeine: 86.5%
  - Exercise: 90.3%
  - Magazine (control group): 87.0%
Among the 3 condition groups, exercise has the highest mean accuracy (90.3%), followed by magazine (87.0%) and caffeine having the lowest mean accuracy (86.5%) among the 3 condition groups. There is a smaller difference (0.5%) between caffeine and magazine groups. These results suggest that exercise is beneficial in promoting better emotional judgment or reasoning and supports the notion that caffeine intake causes people to become more gullible, although only a little bit more than usual. It is worth noting that the 3 conditions are among the top 3 highest in standard deviation for classification accuracy results, with the magazine (control group) having the highest standard deviation (5.3), followed by caffeine (4.7) and then a tie between exercise and male category (4.3). This is depicted in Figure 4.

Figure 4: Standard Deviation of POTUS Classification Accuracy Based on 8 Categories

- Gender
  - Male: 88.3%
  - Female: 87.2%

The mean accuracy difference is 1.1% between male and female participants (male participants performing somewhat better), which suggests that gender does not have a huge effect on the emotional reasoning of a participant.

- Ethnicity
  - Chinese: 88.3%
  - Other: 87.1%
Participants whose ethnicity is Chinese have a slightly higher mean accuracy (88.3%) than those belonging to other ethnicity groups (87.1%), which indicates that Chinese participants seem to be 1.2% more accurate at identifying Trump's lying statements compared to others.

**Figure 5:** POTUS Classification Percentage Accuracy for 20 Iterations Based on 8 Categories

Based on the above classification accuracy results, which are summarized in Figures 5 and 6, the highest mean accuracy is the exercise group with 90.3%, and the lowest figure, with all of the POTUS EDA dataset is 86.4%, followed by the caffeine group (86.5%) with just 0.1% difference. The difference between the highest and lowest mean accuracy among all groups is 3.9%.

**Figure 6:** POTUS Classification Percentage Accuracy for 20 Iterations Based on Statistical Summary
## Table 4: Summary Statistics of POTUS Classification Percentage Accuracy for 20 Iterations Based on 13-sized subsets

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>std</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>94.7</td>
<td>1.4</td>
<td>91.8</td>
<td>96.7</td>
</tr>
<tr>
<td>All Subset 1</td>
<td>86.7</td>
<td>3.5</td>
<td>79.1</td>
<td>93.0</td>
</tr>
<tr>
<td>All Subset 2</td>
<td>86.7</td>
<td>2.7</td>
<td>79.5</td>
<td>90.9</td>
</tr>
<tr>
<td>All Subset 3</td>
<td>85.9</td>
<td>3.0</td>
<td>78.4</td>
<td>91.9</td>
</tr>
<tr>
<td>All Subsets (Avg)</td>
<td>86.4</td>
<td>3.1</td>
<td>79.0</td>
<td>91.9</td>
</tr>
<tr>
<td>Male</td>
<td>91.8</td>
<td>2.3</td>
<td>87.1</td>
<td>95.2</td>
</tr>
<tr>
<td>Male Subset 1</td>
<td>88.0</td>
<td>3.9</td>
<td>81.6</td>
<td>94.7</td>
</tr>
<tr>
<td>Male Subset 2</td>
<td>87.9</td>
<td>4.1</td>
<td>82.1</td>
<td>97.4</td>
</tr>
<tr>
<td>Male Subset 3</td>
<td>89.0</td>
<td>4.9</td>
<td>79.1</td>
<td>97.7</td>
</tr>
<tr>
<td>Male Subsets (Avg)</td>
<td>88.3</td>
<td>4.3</td>
<td>80.9</td>
<td>96.6</td>
</tr>
<tr>
<td>Female</td>
<td>91.4</td>
<td>3.4</td>
<td>85.0</td>
<td>96.7</td>
</tr>
<tr>
<td>Female Subset 1</td>
<td>86.9</td>
<td>4.7</td>
<td>74.4</td>
<td>93.0</td>
</tr>
<tr>
<td>Female Subset 2</td>
<td>84.3</td>
<td>3.4</td>
<td>76.7</td>
<td>88.4</td>
</tr>
<tr>
<td>Female Subset 3</td>
<td>90.3</td>
<td>4.1</td>
<td>83.7</td>
<td>95.3</td>
</tr>
<tr>
<td>Female Subsets (Avg)</td>
<td>87.2</td>
<td>4.1</td>
<td>78.3</td>
<td>92.2</td>
</tr>
<tr>
<td>Chinese</td>
<td>92.2</td>
<td>1.7</td>
<td>89.2</td>
<td>96.4</td>
</tr>
<tr>
<td>Chinese Subset 1</td>
<td>92.6</td>
<td>3.1</td>
<td>86.4</td>
<td>97.7</td>
</tr>
<tr>
<td>Chinese Subset 2</td>
<td>85.2</td>
<td>4.8</td>
<td>79.1</td>
<td>97.7</td>
</tr>
<tr>
<td>Chinese Subset 3</td>
<td>87.2</td>
<td>4.3</td>
<td>79.5</td>
<td>93.2</td>
</tr>
<tr>
<td>Chinese Subsets (Avg)</td>
<td>88.3</td>
<td>4.1</td>
<td>81.7</td>
<td>96.2</td>
</tr>
</tbody>
</table>

As shown in Table 4, the averaged mean accuracy of 3 subsets for all POTUS participants is 86.4%, for male, female and Chinese participants are 88.3%, 87.2% and 88.3%, respectively. Each subset contains 13 participants randomly chosen from its respective dataset segmentation category. In Figure 7, the mean accuracy of all POTUS participants, including Chinese, male, and female category segmentations are displayed next to the mean accuracy of 3 subsets of 13 random participants within the respective categories.
The single factor ANOVA test was implemented across all 8 different categories as well as across all subset categories, returning p values of 8.75142E-06 and 1.55E-09, respectively, indicating that the categories were significantly different from each other and that the results were statistically significant (p < 0.05). Additionally, two-tailed T-test (Paired two sample for means) for all pairs of groups were performed and statistically significant results were found in multiple pairs. The exercise group pairs have shown significant difference with most of the other groups.

From the results discussed above, it is evident that the first hypothesis (H1) was correct, and the remaining two (H2, H3) are partially correct.

H1: Results show that in general, participants who had done the exercising activity performed the best in emotionally recognising or detecting lies from statements. In other words, walking has had a positive effect on affective perception, at least within the observed duration of the two-part experiment that lasted approximately 30 minutes in total.

H2: Participants who had caffeine on average did not do as well as the exercising participants and had only slightly higher mean accuracy than the control group participants, who read a magazine instead. This suggests that caffeine has the least positive effect in affective perception, next to the control condition, and that caffeine tends to make a person more gullible and believing in what others say.
H3: Under normal circumstances in the control activity condition (reading magazine), participants generally perform similarly (only slightly better) in emotional perception tasks in comparison with participants who had drinks containing caffeine. The reading magazine group did not do as well as participants who had exercised, although it should be noted that this group had the most varied results and widest spread of data compared to other groups.

In terms of gender and ethnicity, it is observed that participants who are either male or Chinese had performed better in the affective experiment.

4.2. Pilot study (EmotiW)

The neural network was similarly trained and tested on the pilot study, in which 3 participants watched videos from the EmotiW dataset. All 3 participants have the same gender and ethnicity (Chinese male). Two of them belonged to the reading magazine group for the activity part of the experiment and the remaining one belonged to the exercise group. The classification accuracy results were very low, even lower than chance (50%), with the highest being 42.9% and the mean accuracy was 25.5%. The results are not surprising due to the relatively small sample size of training data compared to having 7 possible classification labels, which are the 7 types of emotions – Happy, Sad, Angry, Fear, Surprise, Neutral, and Disgust. As this was a pilot study with only 3 participants belonging to the same demographic category, future research could explore with a bigger sample size and a more diverse pool of participants.
Chapter 5: Conclusion

According to the classification results in Chapter 5, we have discovered that exercise has the most positive effect whereas caffeine has the least positive effect on affective perception. Additionally, several conclusions can be derived from observing the results. Participants who are either male, of Chinese ethnicity, or both seem more likely to perform better in affective-related tasks, particularly in emotionally recognising the truthfulness of statements. This is an interesting discovery because based on traditional belief and previous studies, females tend to be more emotional, hence performing better in emotional reasoning tasks [12]. Another observation which reinforces Gu's findings [9] is that contrary to common belief, caffeine makes people slightly more gullible compared to usual.

There were some identified limitations in this study, which includes having to stop the collection of data due to the COVID-19 situation and campus closure. The E4 wristwatch and TheEyeTribe eye-tracking device were older models which have limited compatibility and support. Occasionally, TheEyeTribe has difficulty detecting certain eye shapes and this also applies to several participants who wore eyeglasses, although it worked well for others. The 3 missing EDA data were lost and unable to be recovered during the syncing process to the cloud storage. For future works, the use of alternative physiological tracking devices may be considered.

5.1 Future work

Extended research can be done by applying other analysis techniques, such as using Genetic Algorithms [27], on the current data, and using other statistical evaluation measures, such as F1-measure [24] on the current neural network model. Future work also includes an in-depth analysis of other types of physiological signals, namely BVP, HRV, ST, and pupillary response captured by TheEyeTribe, and participant’s facial expressions captured by the webcam. Ideas for future research which were originally brainstormed but not covered within the scope of this project are discussed below.

Other influencing factors that can be considered are experimenting with varying activity durations, the effect of music [23], and other combinations of factors such as music and exercise (both having positive effects), or music and caffeine (positive and negative effects, respectively). It would be interesting to investigate how positive and negative factors in conjunction would influence how well we emotionally perceive. More research can be done to investigate how caffeine intake and walking at certain times of the day or how long the after-
effects of drinking coffee and walking affects our ability to emotional perception. Further investigation could be performed on the EmotiW dataset – the ability in recognising the type of expressed emotions [1], [9] rather than detecting the truthfulness of statements. We may also consider looking into the extent of how biases and preconceptions affect our emotional judgment.

The discovery of factors influencing affective acuity provides many social benefits and contribution to the AC field. These findings provide an indicator of the reliability of future results, depending on whether participants have exercised or drank coffee before doing an emotion-related experiment and depending on their gender and ethnicity background. Unique insights on human creativity and empathy can be derived from these findings to support future HCC research in better ways to build emotionally intelligent systems which can help increase retention, promote creativity, encourage emotional awareness, and improve therapy sessions.
References


### Appendix 1

**Final Project Description Detailing Tasks and Expected Outcomes**

---

**INDEPENDENT STUDY CONTRACT**

*Note: Enrolment is subject to approval by the projects co-ordinator*

**SECTION A (Students and Supervisors)**

<table>
<thead>
<tr>
<th>UnitID:</th>
<th>u6462117</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAMILY NAME:</td>
<td>Xue Ling (Celine)</td>
</tr>
<tr>
<td>PROJECT SUPERVISOR (may be external):</td>
<td>Sabina Caldwell, Jessica Rahman</td>
</tr>
<tr>
<td>COURSE SUPERVISOR (a RSCS academic):</td>
<td>Tom Geddes</td>
</tr>
<tr>
<td>COURSE CODE, TITLE AND UNIT:</td>
<td>COMP4560</td>
</tr>
</tbody>
</table>

| SEMESTER | ☑ S2 YEAR: 2019 | ☑ S1 YEAR: 2020 |
| PROJECT TITLE: | Factors influencing affective acuity |

**LEARNING OBJECTIVES:**
- Experience with affective reasoning
- Experience with neural network applications in physiological signal processing
- Experience with physiological signal tracking devices

**PROJECT DESCRIPTION:**
- Write a literature survey
- Conduct and experiment to test affective perception under variable conditions (such as walking or consuming a beverage or similar) using one or more of:
  - existing HCC data sets on perceived doubt,
  - genuine versus posed emotion or
  - a novel data set
- Record cognitive responses and physiological signals
- Collect and analyse data to determine effects of Variable on affective perception
- Write report

---

Research School of Computer Science  
*Form updated Jul-19*
Appendix 2

Details of Independent Study Contract

<table>
<thead>
<tr>
<th></th>
<th>(fixed)</th>
<th>□ Honours (24 credit)</th>
<th>□ Projects (6-12 credit)</th>
<th>.../ (fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessed project component</td>
<td>% of mark</td>
<td>Thesis</td>
<td>(85%)</td>
<td>Thesis (reviewer mark)</td>
</tr>
<tr>
<td>Presentation</td>
<td>(10%)</td>
<td>Artefact (supervisor project mark)</td>
<td>...45... %</td>
<td></td>
</tr>
<tr>
<td>Critical Feedback</td>
<td>(5%)</td>
<td>Presentation</td>
<td>(10%)</td>
<td></td>
</tr>
</tbody>
</table>

**MEETING DATES (IF KNOWN):**

Weekly

**STUDENT DECLARATION:** I agree to fulfill the above defined contract:

Signature: [Signature] Date: 26 July 2019

**SECTION B (Supervisor):**

I am willing to supervise and support this project. I have checked the student's academic record and believe this student can complete the project.

Signature: [Signature] Date: 26/7/19

**Reviewer:**

Name: [Name] Signature: [Signature]

**Reviewer 2:** (for Honours only)

Name: [Name] Signature: [Signature]

**REQUIRED DEPARTMENT RESOURCES:**

**SECTION C (Honours / Projects coordinator approval)**

Signature: [Signature] Date: [Date]
Appendix 3

3.1 Participant Consent Form

WRITTEN CONSENT by Participants

I have read and understood the Information Sheet you have given me about the research project, and I have had any questions and concerns about the project addressed to my satisfaction (listed here: ).

I agree to participate in the project. YES ☐ NO ☐

Agreement includes: I agree to my participation being audio recorded. YES ☐ NO ☐

And also includes: I agree to my participation being video recorded. YES ☐ NO ☐

I agree to be identified in the following way within research outputs:

Full name YES ☐ NO ☐ we suggest “NO”
Pseudonym YES ☐ NO ☐ we suggest “YES”
No attribution YES ☐ NO ☐ we suggest “YES”

Name: .................................. Signature: .............................................

ANU UniID: ............................ Date: ..................................................

Would you like a summary of the research when it is finished? YES ☐ NO ☐

If YES, please provide an email address: ..................................................

Please list any Special Considerations (e.g. any medical conditions) you have which you would like to bring to the attention of the user study supervisor

☐ Empathy experiment Non-disclosure agreement:

• I agree to not distribute the content of the videos (so will not take photos or videos or audio away from the experiment room).
• I agree to not disclose the specific locations, dates, names of organisations or persons mentioned in the videos, but will refer to them in general terms.
• Signature: ..................................................
3.2 SONA Experiment Description

3.3 Participant Information Sheet

**Participant Information Sheet**

**Project Title:** Human Emotion and Empathy

**Researchers:**
The researchers conducting this experiment are Jessica Sharmin Rahman, Yang Liu, Amelia Poetter, Ruimin Chu, Xue Ling Teh, Sabrina Caldwell and Tom Gedeon.

**Experiment:**
In this experiment, you will watch a selection of videos of varying emotions. Your task is to answer some questions about those videos. In addition, for the first 15 minutes of the experiment you need to complete one of the following tasks:

1. Have a coffee
2. Go for a walk
3. Read a magazine

The experimenter will guide you throughout the process. During the experiment, you will wear a wrist worn device with an aim to collect your Heart Rate Variability (HRV), Blood Volume Pulse (BVP) and Galvanic Skin Response (GSR) and Skin Temperature (ST). Your eye gaze and pupil dilation will be tracked by an eye tracking device placed in front of you. A webcam will be placed in front of you to record your facial expressions.

**Devices**
In this experiment, your physiological signals will be captured by three devices:

1. Empatica E4 device, which is a black roundish watch with a button on it, aiming to collect HRV, BVP and GSR;
2. TheEyeTribe eye-tracking device, which is a black bar tracking eye gaze point and pupil dilation.
3. Logitech C922 Pro Stream Webcam
Tasks
In this experiment, you will watch series of videos on the computer screen and answer questions involving those videos. After completion of tasks, all sensors will be removed. Only during the observation and identification of the experiment, the equipment will be recording your biometrics.

At the start of the experiment, you will:
   i) Strap the device firmly to your non-dominant wrist.
   ii) Press the button on the devices to start recording (the researcher will help you with it).
   iii) Fill demographic information.
   iv) Read instructions and press the Next button.
   v) Watch a series of videos and answer the questions. Press Next to continue.
   vi) Repeat step v) until all the videos have been displayed.
   vii) Fill in the post-experiment questionnaires and press Submit.
   viii) Read the Thank you page and wait patiently for the experimenter to collect the form, and give you any further instructions.

Use of Data and Feedback
The data collected will be used to draw conclusions about certain interaction techniques and the nature of the tasks. Any data collected, either raw or processed, may be used research and publications. The data will be made unidentifiable so that no participant will be able to be identified from any data collected.

Voluntary Participation & Withdrawal
This usability experiment is completely voluntary. You may end the test session or ask for a break at any time. You may request that any or all data collected from you be destroyed. You have the right to completely withdraw from the experiment at any point with no explanation to the researcher. In this case, your data and personal information will be destroyed in accordance with the ANU Code of Research Conduct. You can ask that your name be deleted from our contact list for future testing at any time.

What does participation in the research request of you?
The main purpose of the user study is to collect data to enable useful information to be gained on the interface, the interaction techniques, and tasks. We will give you a pre- and post-task questionnaire that may contain some questions of an identifying nature. You do not need to complete these or any of the other questions if you have any objections to them. The task carried during the session will involve recording of HRV, BVP, GSR, ST, eye gaze and pupil dilation data.

Location and Duration
The study will take place in N239, Level 2, Computer Science and Information Technology (CSIT) Building 108 on the ANU Campus. The time needed to complete this user study will be about 30-60 minutes in one standalone session. This time will include an introduction to the tasks, setup, and completion of the tasks mentioned above.
Incentives
No incentives are provided. Participants signing up via the SONA system gain course credits.

Risks
As the study is conducted in a carefully designed lab environment, all care will be taken to make participants as comfortable as possible, given the nature of the interaction tasks. Some physical discomfort such as wrist and muscle strain may occur with some people including, in rare cases, motion sickness. Participants are free to request that your participation in the user study cease at any stage without explanation.

Confidentiality
The data from the experiment will be made unidentifiable so that no participant will be able to be identified from any data collected. All results published will be in regard to the overall findings from the cohort of participants and not on an individual basis. Until that time, if you give your permission, your contact details will be retained for follow-up testing. The data may be used in follow-up research by researchers not listed on this form. All researchers that will gain access to the data collected in this research will be listed under the same human ethics protocol as the current researcher.

Data Storage
The data from the research will be stored securely at the CSIT Building, ANU. The data from the experiment will be made unidentifiable to retain privacy of each participant. The lookup for the unidentifiable data will be kept in a separate secure location so that participants information can be found in the case of their wanting access to their data or destruction of their data.

In accordance with the ANU Code of Research Conduct all data collected for the research will be stored for at minimum 5 years from the data of publication. After this period the data will be archived for follow-up research. The data will be kept in secure storage at the Research School of Computer Science, ANU.

Queries and Concerns:
If you have any further requests for information or queries regarding the study participants should be directed to the primary investigators,

Jessica Sharmin Rahman
Office: N320, CSIT Building, ANU
Email: jessica.rahman@anu.edu.au

Prof Tom Gedeon
Office: N332 / N331, CSIT Building, ANU
Email: tom@cs.anu.edu.au

Ethics Committee Clearance:
The ethical aspects of this research have been approved by the ANU Human Research Ethics Committee. If you have any concerns or complaints about how this research has been conducted, please contact:

Ethics Manager  
The ANU Human Research Ethics Committee  
The Australian National University  
Telephone: +61 2 6125 3427  
Email: Human.Ethics.Officer@anu.edu.au

3.4 Pre-experiment Questionnaire (Demographic)

Data was collected by a Django program.

Ask subjects:

'name', 'age', 'uid', 'gender', 'glasses', 'ethnicity', 'primary_language', 'subject_source (which course they come from)'

3.5 Questions Related to Videos
a. **POTUS (Detecting Lies) video set:**

**Q1:** How does the information presented in this video look to you?

1 - Completely lying, 2 - A bit misleading, 3 - Don't know, 4 - Probably telling the truth, 5 - Definitely telling the truth

**Q2:** How do you rate your confidence level?

1 - Not confident, 2 - A little confident, 3 - Average confident, 4 - Confident, 5 - Very confident

**Q3:** Have you seen this video before?

1 - Yes, 2 - No

### 3.6 Post-experiment Questionnaire

a. For all experiments:

The ERQ form

![Image of the ERQ form](image-url)

Please answer the questions and click the Submit button.
b. POTUS videos:

   Were any of the videos on topics you are familiar with?

3.7 System Specifications

Software

- MATLAB R2017a
  - Mainly for feature calculation and neural network training
- MATLAB R2019b

System and processor

- Intel(R) Core(TM) i7-8550U processor CPU @ 1.80 GHz, 16.00 GB of RAM and Microsoft Windows 10 Pro 64-bit operating system
  - Mainly for feature calculation and neural network training as it has a faster processing time
- AMD Ryzen 5 2500U with Radeon Vega Mobile Gfx 2.00 GHz, 8.00 GB and 64-bit operating system
Appendix 4

README file

Factors influencing affective acuity
Xue Ling Teh
xueling.teh@anu.edu.au
Last updated: 12 June 2020

The main analysis in the report is based on participants’ recorded EDA physiological signals. Unless specified with _emotiw in the file name or in specific section headings within the file, the files are generally associated with the main POTUS participants.

Source code

[MATLAB]
organize_data: Adds timestamp to the entire EDA data based on the start time and sampling rate (4 Hz/second) recorded on the E4.

normalize_data: normalizes the whole data into 0-1 range, adds results to a new sheet

filter_data: filters noisy data points using median filtering, adds results to a new sheet

merge_data: segments the normalized/filtered data based on video length

sort_data: takes the data from all the participants files and categorizes them based on video fact check ratings

feature_calculation: calculates statistical features

neural_net: classification of fact check rating labels based on statistical features

[Python Jupyter Notebook]
Preprocess Data: Preprocessing and extracting raw data from text files, merging videos watched by all participants into an Excel sheet as reference, merging video start and end timestamps and the fact check ratings for each respective video watched by each participant in separate Excel sheets

Preprocess Physiological Data: Renaming original EDA data file names and moving to a new directory, identifying participant IDs belonging to each activity condition, gender and ethnicity group, segmenting feature files into different subsets