RETRODICTIVE INFERENCES FROM SPONTANEOUS AND SIMULATED FACIAL EXPRESSIONS

Kathleen Kang Yen Li

4231241

Thesis submitted to the University of Nottingham

For the degree of Doctor of Philosophy

September 2017

ABSTRACT

This thesis aimed to investigate whether people can accurately retrodict from a brief sample of others' behaviour to the antecedent event, from spontaneous and simulated expressions. Additionally, the thesis aimed to investigate if retrodiction involves the inference of inner states, unlike a classification task. The experimental procedure enables us to capture a variety of brief, dynamic and spontaneous expressions which closely resembles the expressions seen in real-life social situations. At the same time, it enables us to measure participants' performance against an objective criterion. This paradigm consists of two phases, targets and perceivers. Targets captured in video-clips were either looking at emotional expressions (Experimental Chapter 1 and 2) or pictures from the International Affective Picture System (IAPS) (Experimental Chapter 3 and 4). As a cover activity to provoke spontaneous reactions, targets were told to remember the pictures for a memory test. In Experimental Chapter 3 and 4, they were also asked to simulate expressions which were either congruent or incongruent with what they were viewing. In order to further identify if perceivers were truly inferring targets' inner states, their ("retrodicting perceivers") behavioural responses and eye movements were compared against perceivers who were asked to classify target responses ("classifying perceivers"). The findings revealed that retrodicting perceivers were slightly above chance level in judging the emotional expressions viewed by targets (Experimental Chapter 1 and 2) from spontaneous reactions. As for

simulated expressions, perceivers were also significantly above chance in judging whether targets were displaying an incongruent expression only when targets were looking at a negative picture (Experimental Chapter 3). However, when perceivers were asked to retrodict to the type of picture viewed, they performed significantly above chance only when targets were behaving spontaneously and when they were asked to simulate congruent expressions. In contrast, they were systematically inaccurate when targets were simulating incongruent expressions. Although this may suggest that perceivers were fooled by targets' reactions, perceivers' responses discriminated between occasions when targets were simulating congruent and incongruent expressions. Furthermore, perceivers' responses also discriminated between occasions when targets were behaving spontaneously and when they were asked to simulate congruent expressions. In this respect, perceivers were not entirely fooled by targets' simulations. Finally, there was insufficient evidence to support the hypothesis that retrodicting perceivers engaged in a different strategy than classifying perceivers and hence it is not possible to conclude that retrodicting perceivers were truly inferring targets' inner states.

LIST OF PAPERS

Kang, K., Anthoney, L., & Mitchell, P. (2017). Seven-to 11-year-olds' developing ability to recognize natural facial expressions of basic emotions. *Perception*, *46*(9), 1077-1089.

- Experiment 2:1 in this thesis was based on this manuscript

ACKNOWLEDGEMENTS

There are a number of people without whom this thesis might not have been written, and to whom I am greatly indebted.

Firstly, I would like to thank God for giving me the strength and patience to persevere through my PhD. Next, I would like to express my sincere gratitude to my supervisors, Professor Peter Mitchell and Dr. Peter Chapman for their encouragement and support. Their critical insights and invaluable guidance have contributed immensely to the development of this thesis. I would also like to thank my family for their unwavering love, emotional support and encouragement. I am also grateful to my academic peers who never fail to engage in intellectually stimulating discussions with me. Finally, I would like to thank my friends for their constant support, assistance and for all the fun moments we have had together in the past few years. Thank you.

TABLE OF CONTENTS

ABSTRACT ii

LIST OF PAPERS iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF TABLES xiii

LIST OF FIGURES xviii

CHAPTER 1 History of Face Processing 1

1.1 Facial expressions of emotions 3

1.1.1 Universality of facial expressions of emotions 4

1.1.2 Developmental trajectory of the recognition of facial expressions of emotion 7

1.2 Limitations of emotion recognition studies 10

CHAPTER 2 Inferences of Inner States 18

2.1 Emotion Recognition and inference of inner states 18

2.2 Prediction, Retrodiction and Traditional "Theory of Mind" 20

2.3 Limitations of Traditional "Theory of Mind" tasks. 22

CHAPTER 3 Retrodiction and Facial Expressions 25

3.1 Advantages of retrodiction 29

3.2 Limitations and challenges 32

3.4 AIMS OF THE THESIS 37

CHAPTER 4 Can individuals guess what others were looking at based on brief samples of their behaviour? 40

EXPERIMENT 1 40

4.1 Introduction 40

4.2 Methods 44

4.2.1 Generation of Target Video Clips 44

4.2.2 Editing and Selecting Video Clips 46

4.2.3 Perceiver Phase 48

4.3 Results 50

4.3.1 Were perceivers systematically accurate in guessing what 51

category of facial expressions targets were looking at? 51

4.3.2 Were perceivers' responses affected by what targets 54

were asked to do? 54

4.3.3 Were perceivers' response times affected by the 57

behaviour of targets? 57

4.4 Discussion 58

CHAPTER 5 Differences in processing strategies between retrodiction and classification based on spontaneous facial expressions 61

EXPERIMENT 2:1 61

5.1 Introduction 61

5.2 Methods 66

5.2.1 Generation of Target Video Clips 66

5.2.2 Editing and Selection of Video Clips 69

5.2.3 Perceiver Phase 72

5.3 Results 76

5.3.1 Were perceivers systematically accurate in their responses? 76

5.3.2 Were there any significant group differences in perceivers' judgements between groups? 79

5.3.3 Were there any significant group differences in perceivers' response times? 84

5.4 Discussion 89

EXPERIMENT 2:2 90

5.5 Introduction 90

5.6 Methods 92

5.6.1 Participants 92

5.6.2 Materials and Apparatus 92

5.6.3 Procedure 94

5.7 Results 96

5.7.1 Behavioural data 96

5.7.2 Eye-tracking data 106

5.8 Discussion 116

5.9 General Discussion 117

CHAPTER 6 Can individuals guess if others are simulating their expressions? 120

EXPERIMENT 3 120

6.1 Introduction 120

6.2 Methods 125

6.2.1 Generation of Target Video Clips 125

6.2.2 Editing and Selection of video clips 127

6.2.3. Perceiver Phase 129

6.3 Results 130

6.3.1 Were perceivers systematically accurate in guessing whether targets were responding to an incongruent cue word-picture pairing? 131

6.3.2 Were perceivers' response times significantly different for different types of stimuli? 133

6.3.3 Did perceivers' intensity ratings discriminate between 134

different types of stimuli? 134

6.4 Discussion 136

CHAPTER 7 Can individuals retrodict from others' simulated expressions? 140

EXPERIMENT 4:1 140

7.1 Introduction 140

7.2 Methods 147

7.2.1. Validation of stimuli 147

7.2.1 Perceiver Phase 156

7.3 Results 159

7.3.1 Behavioural results 159

7.4 Discussion 170

EXPERIMENT 4:2 174

7.5 Introduction 174

7.6 Methods 174

7.6.1 Perceiver Phase 174

7.7 Results 176

7.7.1 Behavioural results 176

7.7.2 Eye tracking data 187

7.8 Discussion 205

7.9 General Discussion 208

CHAPTER 8 Discussion 213

8.1 Research Overview 213

8.2 Summary of results 217

8.2.1 Experiment 1- Can individuals guess what others were looking at, based on brief samples of their behaviour? 217

8.2.2 Experiment 2 – Differences in processing strategies between retrodiction and classifying based on facial expressions 219

8.2.3 Experiment 3 – Can individuals guess if others are simulating their expressions? 222

8.2.4 Experiment 4 – Can individuals retrodict from simulated expressions? 224

8.3 Practical and Theoretical Implications 229

8.3.1 Retrodictive Inferences 229

8.3.2 Spontaneous and simulated target expressions 241

8.3.3 Gaze behaviour 243

8.4 Limitations 247

8.5 Directions for future research 251

8.6 Conclusion 255

REFERENCES 257

SUPPLEMENTARY MATERIALS 291

A. The relationship between expressiveness ratings and participants' responses that targets' expressions was negative, for all judges..... 291

B. The relationship between expressiveness ratings and judges' accuracy, for all judges..... 295

LIST OF TABLES

Table 4.1 Confusion matrix of the percentage of correct and incorrect inferences (and data count) in both conditions. 51

Table 4.2 Example confusion matrix based on Wagner (1993)'s scheme. 52

Table 4.3 Mean (and standard deviation) of unbiased hit rates (H_u) and chance probability (P_c) to targets' reactions to each emotional expression. 53

Table 4.4 Mean (and standard deviation) for perceivers' response time in miliseconds (ms) for both target conditions. 57

Table 5.1 The number of unique female and male targets for each expression viewed by targets, separated by expressiveness level. 74

Table 5.2 Mean expressiveness ratings of each type of reaction, as rated by judges in the validation stage. 75

Table 5.3 Confusion matrix depicting the percentage of responses (and data count) for both groups of perceivers. 77

Table 5.4 Classification perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets. 78

Table 5.5 Retrodicting perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets. 78

Table 5.6 The pattern of rank order of responses from the highest Mean H_u to the lowest Mean H_u. 83

Table 5.7 Mean (and standard deviation) response times (in ms) for both perceiver groups. 85

Table 5.8 Number of unique targets for each type of reaction, according to each expressiveness level. 93
Table 5.9 Mean expressiveness ratings for each type of reaction. 94
Table 5.10 Confusion matrix depicting the percentage of responses (and data count) for each group of perceivers. 96
Table 5.11 Classification perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets. 98
Table 5.12 Retrodicting perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets. 98
Table 5.13 Mean (and standard deviation) response times (in ms) for both perceiver groups. 102
Table 5.14 Rank ordering of Reaction times (RT) for the various expressions in both groups, from the fastest (top) to the slowest (bottom) 105
Table 5.15. Mean (and standard deviation) of gaze duration at the eyes for each type of reaction, separated by each perceiver group. 107
Table 5.16 Mean (and standard deviation) of gaze duration at the mouth for each type of reaction, separated by each perceiver group, after data transformation. 111
Table 5.17 Mean (and standard deviation) gaze duration (ms) at the eyes for each perceiver groups, across three time segments. 114
Table 5.18 Mean (and standard deviation) gaze duration (ms) at the eyes for each perceiver groups, across three time segments, after data transformation. 115

Table 6.1 Data count (and percentage) of perceivers' responses. 131
Table 6.2 Mean values of hit rate (raw), false alarm rate (raw), hit rate (z-score), false alarm rate (z-score) and d'. 132
Table 6.3 Mean (and standard deviation) transformed reaction times. 134
Table 6.4 Mean (and standard deviation) of perceivers' intensity ratings for different types of stimuli. 135
Table 7.1 Linear-by-linear association values between expressiveness ratings and participants' response that targets expressions were negative, and their associated p-values for all judges. 149
Table 7.2 Linear-by-linear association values between expressiveness ratings and participants' accurate responses, and their associated p-values for all judges. 152
Table 7.3 Percentage (and data count) of perceivers' responses for each type of reaction.. 159
Table 7.4 Mean hits, false alarms (raw and normalized) and d' values for both perceiver groups. 160
Table 7.5 Mean (and standard deviation) for the proportion of times perceivers stated that targets' expressions were negative, for congruent and incongruent reactions. 161
Table 7.6. Mean (and standard deviation) for the proportion of times perceivers stated that targets' expressions were negative, for spontaneous and congruent reactions. 163

Table 7.7 Mean (and standard deviation) of response times in ms for each type of reaction, for both perceiver groups. 167
Table 7.8 Mean (and standard deviation) of expressiveness ratings (out of 6) for all types of reactions. 168
Table 7.9. Percentage (and data count) of perceivers' responses for each type of reaction. 176
Table 7.10 Mean hits, false alarms (raw and normalized) and *d'* values for both perceiver groups. 177
Table 7.11 Mean (and standard deviation) of the proportion of times perceivers stated that targets' expressions were positive, for congruent and incongruent reactions..... 178
Table 7.12 Mean (and standard deviation) of the proportion of times perceivers stated that targets' expressions were positive, for spontaneous and congruent reactions..... 181
Table 7.13 Mean (and standard deviation) of transformed response times in ms for each type of reaction, for both perceiver groups. 184
Table 7.14 Mean (and standard deviation) of expressiveness ratings for all types of reactions. 185
Table 7.15 Mean (and standard deviation) of average gaze duration, in ms, at the eye region for both studies. 188
Table 7.16 Mean (and standard deviation) of transformed average gaze duration, in ms, at the mouth region for both studies. 190

Table 7.17 Mean (and standard deviation) of average gaze duration, in ms, at the eyes region for both studies. 191
Table 7.18 Mean (and standard deviation) of transformed average gaze duration, in ms, at the mouth region for both studies. 194
Table 7.19 Mean (and standard deviation) of transformed average gaze duration, in ms, at the eyes across 5 seconds (T1-T5) for both perceiver groups. 196
Table 7.20 Pattern of ranking for perceivers' eye movements from highest to lowest. 200
Table 7.21 Mean (and standard deviation) of transformed average gaze duration at the mouth across 5 seconds for both perceiver groups. 201
Table 7.23 Pattern of ranking for perceivers' eye movements at the mouth from highest to lowest. 204

LIST OF FIGURES

Figure 3.1. Indirect and elaborated route of processing..... 26

Figure 3.2. Direct and simple route of processing..... 33

Figure 4.1. Sample screenshot of target reactions in the "memory condition" to (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) fear, (f) sad expressions..... 47

Figure 4.2. Sample screenshot of target reactions in the "imagine condition" (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) fear, (f) sad expressions. 48

Figure 4.3. Emoticons used to represent the six basic emotions: (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) fear and (f) sad. 49

Figure 4.4. Interaction between 'Target Condition" and "Expression Viewed" 55

Figure 5.1 Indirect and elaborated route of processing..... 64

Figure 5.2 Direct and simple route of processing..... 65

Figure 5.3. Emoticons used to represent the emotional categories: (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) sad and (f) neutral. 69

Figure 5.4. The flowchart depicts the selection and validation of stimuli. 71

Figure 5.5. Targets' least expressive reactions to photographs of emotional expressions 73

Figure 5.6. Targets' most expressive reactions to photographs of emotional expressions 74

Figure 5.7. Interaction between “Expression Viewed” and “Group”. 80

Figure 5.8. Interaction between “Expressiveness” and “Expression Viewed”..... 81

Figure 5.9. Interaction between “Expressiveness” and “Group”..... 87

Figure 5.10. Interaction between “Expressiveness” and “Expression viewed”. 88

Figure 5.11. Example of an experimental set-up, as seen by classification perceivers. 95

Figure 5.12. Example of an experimental set-up, as seen by retrodicting perceivers. 96

Figure 5.13. The interaction between the expressions viewed by targets and their expressiveness level. 101

Figure 5.14. Interaction between “Expressions Viewed” and “Group” 104

Figure 5.15. Interaction between “Expressions Viewed” and “Expressiveness”..... 106

Figure 5.16. The interaction between expressions viewed by targets and expressiveness for the average gaze duration (ms) on the eyes region. 110

Figure 5.17. The interaction between expressions viewed by targets and expressiveness for the average gaze duration (ms) on the mouth region..... 113

Figure 6.1. Examples of targets’ reactions: a) spontaneous reaction to negative picture, b) spontaneous reaction to positive picture, c) simulated reaction to a negative cue word-negative picture pairing (congruent negative), d) simulated reaction to a positive cue word-positive picture (congruent

positive) pairing, e) simulated reaction to a negative cue word-positive picture pairing (incongruent positive), and f) simulated reaction to a positive cue word-negative picture pairing (incongruent negative). 128
Figure 6.2. Interaction between "Picture Type" viewed by targets and "Congruency" of cue word with the picture. 136
Figure 7.1. Elaborated and indirect route of processing..... 144
Figure 7.2. Simple and direct route of processing. 145
Figure 7.3. The relationship between expressiveness ratings and participants' responses that targets' expressions was negative, for all judges (see Supplementary Materials A for a larger version). 151
Figure 7.4. The relationship between expressiveness ratings and participants' accurate responses, for all judges (see Supplementary Materials B for a larger version). 154
Figure 7.5. Example stimuli: a) Spontaneous reaction to a negative picture, b) Spontaneous reaction to a positive picture, c) Congruent reaction to a negative picture, d) Congruent reaction to a positive picture, e) Incongruent reaction to a negative picture, f) Incongruent reaction to a positive picture. 157
Figure 7.6. Perceivers' judgments that targets' expressions were negative for simulated positive and simulated negative expressions. 162
Figure 7.7. Perceivers' judgments that targets' expressions were negative for spontaneous and congruent reactions to positive and negative pictures..... 165

Figure 7.8. Retrodicting and classification perceivers' mean d' values for each type of expression. 166
Figure 7.9. Perceivers' intensity ratings for spontaneous, congruent and incongruent reactions to positive and negative pictures. 170
Figure 7.10. Perceivers' responses that targets' expressions were positive for congruent and incongruent reactions to positive and negative pictures. 180
Figure 7.11. Perceivers' responses that targets' expressions were positive for spontaneous and congruent reactions to positive and negative pictures..... 182
Figure 7.12. Perceivers' expressiveness ratings for spontaneous, congruent and incongruent reactions to positive and negative pictures. 187
Figure 7.13. Average gaze duration at the eyes when perceivers were looking at congruent and incongruent reactions to positive and negative pictures...193
Figure 7.14. Average gaze duration at the mouth when perceivers were looking at spontaneous and congruent reactions to negative and positive pictures. 195
Figure 7.15. Retrodicting and classifying perceivers' average gaze duration at the eyes across 5 seconds for both experiments. 198
Figure 7.16. Retrodicting and classifying perceivers' average gaze duration at the mouth across 5 seconds for both experiments. 202
Figure 8.1. This elaborated and indirect form of processing 230
Figure 8.2. Generate-and-test simulationist model of emotion recognition. 232
Figure 8.3. Reverse simulation model and the modification of the reverse simulation model with an "as-if" loop. 233

Figure 8.4. Simple and direct form of processing 236

Figure 8.5. Generate-and-test model of a simple retrodiction model in retrodicting from targets' behaviour to the antecedent event, without inferring targets' inner states. 238

Figure 8.6. The reverse simulation model of retrodictive inferences. 239

CHAPTER 1

History of Face Processing

The face, compared to the rest of the body, offers the most valuable information for the identification of an individual and for social interaction. When we see a face, we need to process two things, (1) that the face belongs to a unique individual, and (2) the emotional content of the facial expression. In this section, we will review the history of face processing research, particularly pertaining to the recognition of emotions from facial expressions.

Early research on face processing placed much emphasis on facial recognition. When people encounter a stranger, they encode the individual's gender and ethnicity. People are remarkably good at inferring whether a particular face belongs to a male or female, even when gender-specifying features (such as hairstyles) were removed (Burton et al., 1993), and under suboptimal processing conditions when faces were blurred. Apart from people's ability to accurately judge one's gender, people are also accurate at inferring the race and ethnicity of others (Blair et al., 2002). But, people do not encounter the same individual in the exact setting with the same viewing angles. Since the appearance of an individual is rarely identical on any two occasions, it is important to be able to recognize an individual at a later point in time and in a different context. As such, many studies have focused on how people recognise faces, and the differences, if any, in recognising familiar and unfamiliar faces. In a pioneering study by Patterson and Baddeley (1977), participants were required to learn names for unfamiliar faces

and then identify these individuals from a series of photographs. When participants viewed photographs that reflected changes in pose and simple changes in appearances, participants' recognition of unfamiliar faces were compromised. Baddeley and Woodhead (1981) also found a reduced recognition of unfamiliar faces due to changes in head angles. According to Ellis et al. (1979), repeated interactions with a face may lead to the establishment of a structural code which emphasizes the facial features, highlighting the importance of the role of familiarity in face recognition. Overall, familiar faces were recognized more quickly and accurately as compared to unfamiliar faces. Thus, while people are accurate at forming various inferences about one's identity, the ability to recognise a face depends on the familiarity of the face.

Following on from the importance of familiarity in face recognition, Bruce and Young (1986) developed a theoretical framework of face recognition which holds a prominent position in face processing research until today. In this framework, they proposed that there are seven types of information which can be derived from faces, that is, pictorial, structural, visually derived semantics, identity-specific semantics, name, expression and facial speech codes. This model states that the recognition of faces begins with the perceptual processing of the facial features and their configural relations, prior to progressing through specialized functional modules which are responsible for recognizing specific types of information in the face. This model assumes that facial expressions are not important for face recognition. Since the recognition of facial expressions is sub-served by a distinct process from face recognition, judgements about facial expressions are not affected by familiarity (Young et al., 1986). Drawing upon the notion that facial expressions and face recognition may be sub-served by separate

processes, it has been suggested that this dissociation begins at either the structural encoding stage (Calder et al., 2001) or at a recognition stage (Adolphs et al., 1994). This distinction in processing facial identity and facial expressions is supported by studies investigating the role of information concerning facial features and its relative configuration. While the recognition of facial identity is holistic (Tanaka and Farah, 1993) and depends on configural information about spatial relations among features (Moscovitch et al., 1997), there have been mixed findings pertaining to the recognition of facial expressions with support for both feature-based processing (Cottrell et al., 2001) and configural processing (Calder et al., 2000). Therefore, past research on face processing has established a distinction between the processing of facial identity and facial expressions. Although it is important for us to recognize the identity of others, it is also crucial for us to be able to recognize the facial expressions of others as it sets the tone for social interactions (amongst other things).

1.1 FACIAL EXPRESSIONS OF EMOTIONS

Faces can provide a wealth of information regarding the gender, race, familiarity of an individual, as well as information about the emotions of others. The ability to recognise emotional cues from facial expressions is necessary for social interaction. Darwin (1872) suggested that facial expressions are the universal language to signal emotional states. Since then, many researchers have often associated emotional states with facial expressions. Woodworth (1938) suggested that faces convey families of emotions, if not specific emotions. Later researchers identified a prototypical expression associated with each type of

emotion, for a set of discrete emotions (Ekman, 1972; Izard, 1971, 1977; Tomkins, 1962). Based on these propositions that facial expressions convey emotions, many studies have focused extensively on whether there is a universal recognition of a certain set of emotions and have questioned the form of the developmental trajectory in the ability to recognise facial expressions.

1.1.1 Universality of facial expressions of emotions

In the 1970s, research on facial expressions had mainly focused on the recognition of basic emotions. Basic emotions are associated with a cluster of six (happy, surprised, sad, angry, disgusted, fear) expressions, for which a specific prototype expression can be identified. The proposition that there is a set of six basic emotions comes from findings suggesting that there is a set of emotions which are expressed and recognised in the same way across different cultures (Ekman, 1992). As a way of testing people's ability to recognise emotions, posed and static facial expressions are often shown to observers (henceforth, perceivers) who are asked to judge what emotion is shown by labelling the expressions or categorizing them according to the appropriate label. If perceivers in different cultures label or categorize the expressions with the same term, then this is interpreted as evidence of universality. Perhaps, the most famous study which supports this concept of universal basic emotions is Ekman, Sorenson and Friesen (1969)'s study in pre-literate and literate cultures. Using an alternative forced-choice task, they found that there was a high agreement in the accuracy of judgements of a set of standardized photographs depicting the facial expressions of six primary emotions across both pre-literate and literate cultures (New Guinea, Borneo, United States, Brazil, Japan). These findings were also supported by Ekman and Friesen's (1971) study which found that members of the Fore group, who were preliterate

tribesmen of Papua New Guinea, associated facial configurations with discrete emotions which corresponded to the six basic emotions. These two studies, therefore, support the universality of recognising a certain group of facial expressions of emotions among pre-literate and literate cultures.

Apart from the universality of facial expressions of emotions in pre-literate and literate cultures, some studies also found that the recognition of emotions was universal in different literate cultures. In a cross-cultural study where Americans, Malaysians and Temuan aborigines were asked to judge whether a photograph of facial expressions corresponded to the six basic emotions, both Americans and Malaysians accurately identified the expressions (Boucher and Carlson, 1980; McAndrew, 1986). In another study where Chinese participants were asked to label 3D-morphed emotional faces which were developed along a happiness-surprised-fear-sadness-disgusted-anger-happy continuum, Wang et al. (2006) found that Chinese participants were accurate in labelling each emotion, consistent with the categorical perception of the six basic emotions. Not only is there a universal recognition of six basic emotions, contempt was also found to be universally recognised. When participants of different cultures viewed a standardized set of photographs from the Japanese and Caucasian Facial Expressions of Emotions (JACFEE) and selected a single emotional term from seven choices, Biehl et al. (1977) found a high agreement in identifying happiness, surprised, sadness, anger, disgusted, fear and contempt between participants from Hungary, Japan, Poland, Sumatra, United States and Vietnam. In Ekman and Friesen (1986)'s study where participants had to select an emotion term from a list of seven emotion terms (happiness, surprised, sadness, anger, disgusted, fear, contempt), they found that contempt, in addition to the six basic emotions, was also recognized by people

across different cultures such as Estonia S.S.R., Germany, Greece, Hong Kong, Italy, Japan, Scotland, Turkey, United States and West Sumatra. In a similar study by Ekman and Heider (1988), participants from West Sumatra judged photographs of Japanese and American faces with a very high agreement rate on which expressions signalled contempt, in addition to the six basic emotions. A meta-analysis conducted by Elfenbein and Ambady (2002) also showed that emotions were universally recognized at better-than-chance levels. Hence, these studies have shown that there are approximately six or seven emotions which are universally recognised across different cultures.

However, this notion of universality has been challenged by other studies that found accurate recognition of fewer than six emotions cross-culturally. When American and Japanese participants viewed facial expressions of emotions which contained only facial movement that is related to a particular emotion, they were only good at recognizing four out of six emotions (Matsumoto, 1992). Moreover, Jack et al. (2009) showed that Eastern observers used a culture-specific decoding strategy which is inadequate to reliably differentiate between facial expressions of fear and disgusted. In another meta-analysis conducted by Nelson and Russell (2013), there was a large percentage of observers who accurately matched happiness to the predicted facial expression across different cultures. But, for negative emotions, Western participants gave the highest matching scores, followed by Non-Western literate participants and Western illiterate participants, hence, contradicting the claim that people universally recognise basic emotions from signals in the face. Additionally, when Westerners and East Asians were asked to categorize animations of facial expressions according to emotions, there was no universality. While Westerners represented each of the six basic emotions with a

distinct set of facial movements common to the emotional category, Easterners did not. This notion of universality has also been challenged by studies which compared the performance between pre-literate and literate cultures. In Elfenbein et al. (2007)'s study which compared expressions between Quebec and Gabon, they demonstrated evidence for cultural variations in expressive displays for most emotional states such as anger, sadness, surprised, happiness, contempt, shame and serenity. But, there was no cultural variation for fear and disgusted. From Gendron et al. (2014)'s study who instructed participants from the United States and the Himba ethnic group to sort the images of posed expressions into piles by emotion type, the pattern of American participants' responses corresponded with the presumed universal features while Himba participants did not conform to this pattern. Could these cultural differences in recognition be due to the difference in how emotions are conveyed in faces across different cultures? From segregating out the latent expressive patterns of emotional facial expressions across Western and Eastern cultures, there were four latent expressive patterns instead of six which corresponded to the six basic emotions (Jack et al., 2012). These findings thus suggest that the recognition and display of facial expressions of emotions may not be universal as suggested by earlier studies, but may be influenced by cultural factors.

1.1.2 Developmental trajectory of the recognition of facial expressions of emotion

Despite the ongoing debate surrounding the universality of facial expressions of emotions, people do possess the ability to recognise emotions from facial expressions. But, this ability to recognise emotions has different developmental trajectories, depending on the emotion. Similar to the paradigms

used to investigate the universality of emotions, the tasks used in developmental studies often involve matching and labelling. These tasks require participants to either match emotional expressions from an array, or label emotional expressions in a forced-choice paradigm.

Developmental studies have shown that at a very young age, infants can discriminate facial expressions of emotions. Infants aged four to six months old could recognise and discriminate facial expressions of anger, fear and surprised (Serrano et al., 1992). Other developmental studies have investigated the ability to recognise emotional expressions from five years of age up to adulthood. Happiness is the first expression to be accurately recognized and children as young as five can recognise happiness as accurately as adults (Gao and Maurer, 2009; Herba and Phillips, 2004). As for surprised, children aged 5 to 6 are less accurate than older children and adults (Gao and Maurer, 2010; Montirosso et al., 2010), but children as young as 5 can recognize sadness as accurately as adults (Gao and Maurer, 2009). The recognition of anger consistently improves from school age to late adolescence (Montirosso et al., 2010) while there seems to be no improvement between pre-school and adolescence for the recognition of disgusted. The recognition of fear is the most difficult in children (Herba and Phillips, 2004; Widen, 2013). Thus, at the age of five, children can recognize happiness and sadness as accurately as adults, but the recognition of surprised, disgusted, anger and fear was comparably less accurate than adults. Moreover, the recognition of happiness is the earliest, followed by sadness, anger, surprised and disgusted (Herba and Philips, 2004; Widen, 2013), with mixed findings for the recognition of fear.

Nevertheless, more recent studies have shown a different developmental trajectory for the recognition of emotions. In a study by Lawrence et al. (2015) who tested 478 children aged six to sixteen-years-old on standardized, prototypical photographs of the six basic emotions from Ekman and Friesen's Picture of Facial Affect, they discovered significant linear trends in children's ability to recognise facial expressions of happiness, surprised, fear and disgusted. But, there is little or no change in accuracy in recognising angry and sad expressions, implying that pre-schoolers are as accurate as adolescents. In Rodger et al. (2015)'s study which utilized a threshold-seeking algorithm to manipulate the quantity of signals in faces displaying the six basic emotional expressions (and neutral), happiness was the most easily recognized whereas fear was the most difficult. They also concluded that there are three characteristics with regards to the recognition trajectories of six basic emotions. Firstly, disgusted, neutral and anger are expressions that show steep improvements with age. Secondly, there was a gradual improvement with age for the recognition of sadness and surprised. Lastly, the recognition of happiness and fear was already matured by five. Although there seems to be a consensus that happiness is recognized earliest and the recognition of surprised improves with age, the findings pertaining to the recognition of the other emotions are mixed. These findings are in line with adult studies which demonstrated that happiness was recognised more accurately, earlier and faster, while negative emotions were recognised less accurately, latest and slowest (Calvo and Lundqvist, 2008; Tottenham et al., 2009). Thus, children are able to recognise the six basic emotions from facial expressions and this ability improves with age.

Past studies on the recognition of facial expressions of emotions have, therefore, provided us with great insight into people's ability to recognise facial

expressions of emotions across different cultures and ages. However, there are mixed findings regarding the universality of emotions and the developmental trajectory of the recognition of emotions which could be attributed to the different types of stimuli used. In saying this, the stimuli and paradigm used in past studies might not be ideal and these limitations should be taken into consideration for future studies.

1.2 LIMITATIONS OF EMOTION RECOGNITION STUDIES

Although these emotion recognition tasks have demonstrated that people can recognise emotions from displays of facial expressions, there are several limitations surrounding these tasks, in relation to the artificial nature of the stimuli and the lack of an objective answer. The stimuli used in emotion recognition tasks are often static, posed, exaggerated and controlled, such that they do not resemble the dynamic and naturalistic spontaneous expressions seen in everyday social interactions. In Ekman's (1969) study, photographs were selected to obtain only those which showed the pure display of a single affect. A few years later, Ekman and Friesen (1976) developed the Pictures of Facial Affect (1976) which is perhaps, the most widely used and validated series of photographs used in facial expression research. It consists of static photographs of facial expressions of specific emotions. Furthermore, in Ekman and Friesen (1986)'s study, the stimuli were created by asking actors (henceforth, targets) to move certain muscles which were then scored using the Facial Action Coding System (FACS) according to the appropriate emotion category. In addition to Ekman's database of facial expressions, many research groups have followed in their footsteps of creating their own database of

static and posed facial expressions such as the Karolinska Directed Emotional Faces (Lundqvist, 1998), NimStim (Tottenham et al., 2009) and more recently, the Chicago Faces Database (Ma et al., 2015). Despite the ongoing discussions on the importance of ecological validity in experimental research (Krumhuber et al., 2013), many still use these artificial stimuli for their research. The findings from research using these artificial stimuli might give a misleading impression of people's ability to recognise the emotions or social cues in everyday social situations.

Firstly, these expressions are static, unlike the dynamic nature of everyday experiences. Static expressions usually capture the apex of the expression and this static presentation of the apex is displayed for a longer period than the actual duration of the apex in real life. Hence, when participants are exposed to the apex of the expression for a longer period of time, this could give them a stronger signal about the emotion of others, leading to more accurate judgements for static expressions. Nevertheless, the sequence of movements that make up an expression in a dynamic presentation could be uniquely informative temporally (Ekman and Friesen, 1982; Hess and Kleck, 1990), giving an advantage over static displays that over-expose the apex. Thus, the use of static expressions may over-estimate or under-estimate people's ability to recognise social cues in daily lives, leading to a misrepresentation of people's ability to recognise facial expressions of emotions.

Within the face processing literature, there has been mixed findings regarding the contribution of dynamic information to face recognition abilities. Christie and Bruce (1998) found that motion per se is not a crucial factor in establishing representation for unfamiliar faces. Other studies, however, have

shown that there might be an advantage for dynamic information under more difficult circumstances. For instance, seeing a moving face is more advantageous when recognising degraded famous faces but only when the original dynamic characteristics of motion were maintained (Lander and Bruce, 2000). In addition, Thornton and Kourtzi (2002) found that dynamic advantage arises only when the task places more demands on visual working memory. Apart from that, Hill and Johnston (2001) also found that different types of movements are advantageous for different types of categorization. In particular, rigid head movements are useful for distinguishing between individuals while nonrigid motion are useful for categorizing on the basis of sex. Thus, it is suggested that there is an advantage of motion and dynamic information only under suboptimal conditions.

Likewise, relating to the recognition of emotions, there was an advantage of dynamic information, depending on the type and intensity of the expression. Harwood, Hall and Shinkfield (1999) found that dynamic information only, improved the perception of sad and angry facial expressions. In Kamachi et al. (2013)'s study, they discovered that sadness was more accurately described from slow sequences. Accuracy for happiness was better from faster sequences while accuracy for anger was better for medium sequences. However, when subtle facial expressions were used, there was a highly robust effect of motion across all six basic emotions (Ambadar et al., 2003). Not only that, there is an advantage for dynamic information for different populations. While there was no advantage for dynamic over static presentation of expressions for controls, slow and dynamic presentation of facial expressions facilitates facial expression recognition by autistic children. Therefore, the role of dynamic information in the recognition of emotional expressions is not straightforward but depends on many factors. As

such, while people can recognise static facial expressions, the recognition of dynamic facial expressions in daily life is dependent upon many other factors.

Apart from the static nature of these stimuli, the facial expressions used are often posed and exaggerated and are, therefore of lower ecological validity. Specifically, spontaneous expressions differ from posed expressions in terms of appearance and timing. For example, spontaneous smiles are smaller in amplitude, longer in duration, and slower in onset and offset times, in comparison with posed smiles (Ekman, 2003). When perceivers were asked to label spontaneous expressions of targets responding to emotion-eliciting slides, recognition was above chance for happiness, disgusted and anger, but not for sadness, fear, neutral and surprised (Wagner et al., 1986). Similar findings were discovered by Hess and Blairy (2001) who found that observers identified happiness more frequently, followed by sadness, anger and disgusted. Additionally, Motley and Camden (1988) also discovered that only spontaneous expressions of happiness were recognised with any level of accuracy and perceivers' recognition of emotion was significantly higher for posed than spontaneous expressions. But, a recent study by Sauter and Fischer (2017) showed that people reliably recognised emotions from both posed and spontaneous expressions, from facial expressions of various databases. But, the question as to whether posed expressions were recognised more accurately than spontaneous expressions largely depended on the stimulus set used. Hence, while people can recognise emotions from posed expressions, most studies showed that people are only accurate at recognising happy expressions from spontaneous facial expressions. Moreover, the exaggerated nature of the stimuli also makes it difficult for perceivers to misclassify the facial expressions since there is a strong

signal, hence, potentially over-estimating people's ability to recognise emotions in everyday social situations.

In published research, facial expressions of emotions are often created under controlled environmental conditions where the lighting, position of the face relative to the screen and colour properties of photographs are constant. Expressions are also often selected on the basis that they resemble a prototypical expression (Ekman, 1969). However, everyday expressions can be variable in terms of expressiveness. As one particular emotion may be displayed in many different ways and in varying intensities, there is a huge range of within-person variability for any given emotion. As such, the stringent process which is often used in selecting facial expressions gives rise to an artificial lack of variability in the stimuli used. In recent years, this issue pertaining to the lack of variability has been prominent in discussions of face processing research (Burton et al., 2011; Burton, 2013; Burton et al., 2016; Jenkins et al., 2011). Photographs of faces are not consistent indicators of facial appearance because they often ignore within-person variability (Jenkins et al., 2011). Burton (2013) emphasized that the use of artificial or tightly controlled stimuli prevents the generalisability of results to naturally occurring stimuli. As real faces vary in terms of their properties, these artificial and controlled stimuli give false prominence to certain characteristics of faces. By creating artificial and controlled facial expressions, there is inappropriate emphasis on the importance of certain facial features for identifying emotions. People vary in systematic ways but the variability in one face does not generalise well to variability in other faces (Burton et al., 2016). To illustrate, a certain emotion can be displayed in many different ways for an individual. But, this variation in how this emotion is displayed is different for each individual;

therefore, it is not something that can be created in a systematic manner within an experimental setting. The benefits of using stimuli with high variability have been demonstrated for face learning. When participants were presented with name and face associations for unfamiliar identities in high and low within-person variability, participants were faster and more accurate on the face matching task for identities learned in high variability as compared to identities learned in low variability conditions (Ritchie and Burton, 2017). Thus, it is important to include a range of expressions with high variability in order to assess people's ability to understand others' expressions in everyday social interactions.

Apart from the limitations pertaining to the artificial nature of the stimuli, another critical limitation is that these tasks lack an objective answer that we can use to evaluate perceivers' responses. Perceivers were often asked to guess targets' emotions from posed photographs of facial expressions. It is often assumed that if perceivers responded with the corresponding emotional category, they are accurate at inferring targets' inner state. But, there is a fundamental issue underlying this assumption. Because these expressions are posed, the expressions that are displayed may not necessarily reflect the true inner state of targets even if there is a consensus that the expression represents the particular inner state category. Posed expressions are not the product of a naturally occurring inner state and actors might even intentionally dissociate from their internal states (Ekman et al., 1988). Under naturalistic settings where targets experience intense emotions, they often produce only subtle cues (Zaki et al., 2009). As such, it cannot be assumed that people's attribution of inner states is accurate on the basis that the displayed expression matches the corresponding emotional category that the target had been instructed to pose. For instance, there is a general consensus that

a smile represents a happy emotion. However, it does not necessarily mean that targets are actually feeling happy if they smiled. If there is no objective answer as to the true inner state of targets, we cannot measure whether perceivers are accurate at inferring others' inner states. In fact, this problem of satisfying the truth condition is one of the major challenges in this field of social cognition research (West and Kenny, 2011). In order to study accuracy, it is important to first define and measure the true inner state that the perceiver is attempting to identify. An objective criterion is important as the goal of everyday social cognition is not to simply draw any type of inference about others, but to use these accurate inferences to guide our social behaviour in response to these inferences.

One method proposed to overcome this problem is the empathic accuracy paradigm (Ickes, 1997; Ickes et al., 1990). This method examines interpersonal consensus about internal states where perceivers' ratings of targets' thoughts and feelings are compared with targets' reports on their own states. For instance, in a dyadic interaction, two people rate how they felt during specific times in the interaction. After that, they were asked to make judgements of how they believed their partner felt during corresponding times in the interaction. The truth criterion in this paradigm refers to the target's declaration of his or her feelings. However, this is based on the assumption that targets are stating the truth about their feelings. We can never really be certain the individuals are stating the truth because their mind might not be transparent to itself. In order for targets to know what is in their own mind, they have to undergo a process of introspection. After this process of identifying their inner states through introspection, they have to be sufficiently articulate to be able to translate their inner states into a verbal description that describes exactly the state in question. Then, targets have to

choose to be honest in reporting their inner states. Therefore, we can never really be certain that targets' self-reports of their emotions are what they are truly feeling. Suppose that targets were incapable of identifying their true inner state, they would presumably be viewing themselves from the perspective of a third party and infer the inner states from clues in their own observable behaviour. In such cases, therefore, it is hardly surprising to find an agreement between targets' declaration of their own inner state with perceivers' judgements. As such, Ickes' empathic accuracy paradigm does not satisfy the truth condition. It is important to create a paradigm that is able to fulfil this truth condition in order to accurately measure people's ability to infer others' inner states from facial expressions.

In summary, these limitations of past emotion recognition tasks pertaining to the nature of the stimuli and the problem of an objective answer need to be taken into consideration in future studies to develop an experimental paradigm that captures naturalistic stimuli with an objective answer – only then are we able to paint a true representation of people's ability to read social cues in everyday interactions.

CHAPTER 2

Inferences of Inner States

2.1 EMOTION RECOGNITION AND INFERENCE OF INNER STATES

Despite the limitations of these emotion recognition tasks, studies have shown that people can consistently and reliably label facial expressions using terms like happy, sad, angry, fearful, surprised and disgusted. This finding, however, raises the question of whether people are actually inferring the inner states of others when they label the emotions. In other words, upon recognising an emotional expression, do perceivers infer the inner states of targets or do they merely label and classify their expressions according to the responses? For example, upon viewing a smile, do perceivers infer that the target is feeling happy or do they merely classify it as a happy expression? It should be noted that inferring that the target is feeling happy is different from recognising that the targets' expression is a happy expression. The former requires perceivers, at the very least, to accept that the target has inner states. According to Baron-Cohen et al. (1996), basic emotions are identifiable from facial features alone, thus, the recognition of basic emotions may not require any inference of inner states. But emotions such as interested, jealousy and flirtatious might not be associated with any obvious facial characteristics (Premack and Woodruff, 1978) and therefore, a correct answer requires an inference of inner states. In Baron-Cohen et al.'s study, participants were asked to view paintings of people displaying complex emotions and subsequently were asked to select an appropriate label from a set of alternatives.

Participants were better than chance in their responses, thus leading to the conclusion that facial expressions provide sufficient information for perceivers to infer underlying inner states even when there are no obvious facial characteristics.

Contrary to Baron-Cohen et al.'s opinion, one might suppose that the mind is embodied in one's expressions such that it can be directly perceived from one's expressions, without the need to draw any kind of inner states inference (Gallagher et al., 2008; Zahavi and Gallagher, 2008). The suggestion here is that the recognition of emotions, including those that are basic and those that are not basic, can be achieved by perception alone. Computer modelling studies have demonstrated that the structure present in images of facial expressions is sufficient to generate some structure of emotion categories as perceived by humans (Calder et al., 2000). Moreover, neural network models can also be trained to classify faces for basic (Cottrell et al., 1998; Padgett and Cottrell, 1998) and complex emotions (Kaliouby and Robinson, 2005). It seems that the categorical perception of facial expressions in network models relies predominantly on perceptual information because they do not have the capability to infer or think about the inner states of others.

Since it is possible to classify emotions on the basis of perceptual information alone, it could be the case that people, like neural networks, rely predominantly on the perceptual characteristics in recognising emotions without having to infer the inner states of others. Previous research often used paradigms which require one to discriminate between different emotions, sort different emotions into categories, select prototypes within a category and label emotions in forced choice matching tasks. For these tasks which do not require an explicit

reasoning or inference about the targets' inner states, it would be reasonable to explain perceivers' performance on the basis of a fast, automatic and highly stimulus-driven perceptual capacity (Scholl and Tremoulet, 2000). Upon viewing targets' expressions, perceivers do not have to start thinking about what might be going on in the target's mind because everything that the perceiver needs to know about the target's mind is gained by merely perceiving the target's expression. Since there is already so much information about one's inner states available in the face that perceivers can grasp everything that they need to know in understanding the expression (Ratcliffe, 2007), perceivers can recognise emotions without the need to make any inferences. Thus, asking people to recognise emotions from facial expressions might not be a valid test to investigate people's ability to infer others' true underlying inner states.

2.2 PREDICTION, RETRODICTION AND TRADITIONAL "THEORY OF MIND"

Are there any scenarios in which we make inferences about targets' inner states based on aspects of targets' behaviour that is not expressions of emotion? One such scenario is incorporated in the test of false belief, one of the first tasks developed for use in traditional "Theory of Mind" research. Perceivers are generally asked to make a predictive inference about the behaviour of targets based on what is known of their inner states, namely their beliefs. In the standard false belief task, also known as the change-of-locations task (Wimmer and Perner, 1983), the protagonist, Maxi, puts his chocolate in a green cupboard and then leaves the scene. His mother then enters the scene and in Maxi's absence, moves the chocolate to the blue cupboard. The child participant is then asked the false

belief question which requires the child to predict where Maxi will look for his chocolate ("Where will Maxi look for his chocolate when he returns?"). In order to correctly answer this question, the child participant has to infer Maxi's false belief, that is, he thinks the chocolate is in the green cupboard and thus, predicts that Maxi will look for the chocolate in the green cupboard. If the child answers correctly by predicting that Maxi will look for the chocolate in the green cupboard, then the child is regarded as being able to make an inference of inner sates. On the other hand, if the child makes an incorrect judgement that Maxi will look for the chocolate in the blue cupboard, then it is assumed that he or she failed to infer the inner state of Maxi (i.e., the false belief that the chocolate is in the green cupboard where he last saw it). Children at the age of four years old and above systematically pass this task (Wimmer and Perner, 1983; Wellman et al., 2001). As an accurate prediction of where the protagonist looks for the object is only possible if perceivers accurately infer that the protagonist held a false belief, this task is therefore an example of a task that requires the inference of inner states in the absence of needing to interpret an emotional expression.Because the participants in the false belief tasks were never asked explicitly to reason about the protagonist's inner states when predicting targets' behaviour, it is not known if participants were really inferring the protagonist's inner states.

False belief tasks are not only used to predict the protagonist's behaviour from inner states, but are also used to test perceivers' ability to retrodict from the protagonist's behaviour to his or her inner states. Although the term "retrodict" was introduced by Gallese and Goldman (1998), this concept was previously investigated using false belief tasks. In a retrodiction false belief task, participants are asked to make a backwards inference from a protagonist's behaviour to the

inner state. For example, in Bartsch and Wellman (1989)'s study, children were asked to explain the protagonist's actions in terms of their desires and beliefs. For instance, children were asked to explain why they think a protagonist is eating a rock that looks like a peanut. Three-year-olds were surprisingly good at retrodicting from the behaviour of the protagonist to the protagonist's inner states (in this case, beliefs) despite failing at prediction tasks. Robinson and Mitchell (1995) extended this study by testing children's ability to explain why the protagonist was searching in the wrong place with reference to a more distal cause such as the event that caused the inner state. For instance, in order to infer why the protagonist was searching in the wrong place, children had to refer to the fact that the protagonist had been absent when the object was moved from an initial location to its current location. In such cases, retrodiction refers to the process of inferring the cause of the behaviour, where the cause identified can be proximal (such as beliefs or false beliefs) or distal (an external event such as the object being moved from one place to another in the protagonist's absence). Thus, retrodiction to a proximal or distal cause requires perceivers to make inferences about inner states. Although false belief tasks can be used to predict one's behaviour from one's inner states or retrodict from one's behaviour to one's inner states, there are nevertheless several limitations of this task.

2.3 LIMITATIONS OF TRADITIONAL "THEORY OF MIND" TASKS.

Firstly, the procedure in false belief research utilises artificial stimuli which does not capture the features of real-life social scenarios. Participants are often required to reason about an inner state in a certain scenario which contains all of

the elements decided upon by a researcher. These tasks consist of a narrative about a fictional character and the sequence of behaviours performed in a situation which leaves them in a certain belief state. If one performs accurately on these tasks, then one is said to have acquired an ability to infer inner states. However, the demands faced by people in everyday tasks are very different from the demands that are incorporated in these tasks. In everyday social situations, people have to make inferences from subtle, brief and fleeting cues in facial expressions. Thus, the demands in false belief tasks do not necessarily map onto the demands that we see in everyday social interactions.

Apart from that, another limitation of these tasks is that it does not satisfy the truth condition (West and Kenny, 2011). In false belief tasks, the target is a hypothetical and artificial entity where internal states do not originate within the target, rather, these states are defined by the researcher. As there is no real human target, there is no prospect of satisfying the truth condition because an inner state cannot be created by an artificial target. These tasks rely on the assumption that people react to a particular situation in a similar way. In other words, it is assumed that the protagonist will hold a false belief about the location of an object which will trigger the targets' behaviour to act accordingly, when an object is moved to another place in the absence of a protagonist. Although the assumption that one has a false belief when we see someone looking for an object may seem reasonable, we cannot know for certain that the protagonist truly holds a false belief merely based on the assumption that they did not see the object being displaced. In addition, even if one holds a false belief about an object being displaced, they may not act accordingly, that is, to look at the location where they last saw it. As such, this task does not satisfy the truth condition because we can never really be sure

about the hypothetical protagonist's inner state (i.e. belief)—because the inner state in this case is not a naturally occurring phenomenon.

Furthermore, this task has no scope for taking into account individual differences in targets' behaviour. Although false belief tests probe understanding of the relationship between situations and behaviour, it does not capture the manner in which people normally perform when predicting and explaining others' behaviour in real life. In false belief tasks, participants might judge that the protagonist will look in the current location of the object because they assume that the protagonist is intelligent enough to infer that the object moved from its previous location to a new location (Rai and Mitchell, 2004), not because they lack an understanding of the protagonist's inner states. In relation to this, people often expect others to behave in a variety of ways, depending on their personality and the context of the interaction, suggesting that perceivers often taken into account the traits of others and the situation when explaining the behaviour of others. Thus, these false belief tasks do not capture the variety of ways in which one could react according to individual differences. In summary, the limitations of these tasks suggest that false belief tasks are not suitable to investigate the types of inferences made in everyday social situations.

CHAPTER 3

Retrodiction and Facial Expressions

Rather than inferring others' false beliefs from hypothetical targets, people make retrodictive inferences from facial expressions of others. It is more common to process the facial expressions of others together with other aspects of their behaviour. Although 'retrodiction' is a term only briefly mentioned by Gallese and Goldman (1998) that refers to the backwards inference from targets' behaviour to inner states, this concept can be extended by defining retrodiction as the backwards inference from targets' behaviour to an external event that happened in the world, where the proximal cause is the inner state caused by the event and the distal cause is the external event which caused the inner state and subsequently, the behaviour of targets (Teoh et al., 2017) (Figure 1). This process does not only involve labelling the expression or inner states, but requires the perceiver to infer the event preceding the inner state which, in turn, caused the behaviour in the target.

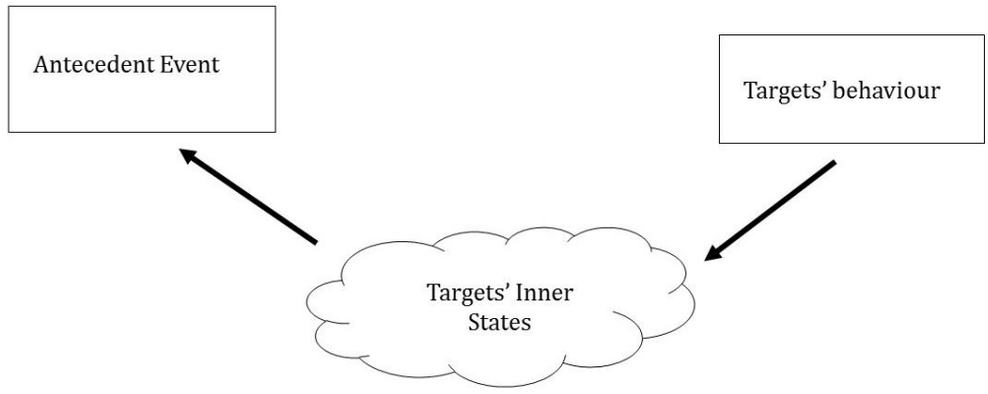


Figure 3.1. Indirect and elaborated route of processing. Upon viewing targets' behaviour (represented at the right of the figure), perceivers make a backwards inference from the behaviour to the proximal cause of the behaviour which is the inner state of targets. Then, drawing upon this inference of targets' inner states, perceivers make a further backwards inference to the distal cause of the behaviour, i.e. the possible antecedent event of the inner state.

Because we live in a complex social environment, this definition of retrodiction considers how the inference of inner states is an adaptive process which equips the perceiver not only with knowledge of the target's inner states but also knowledge of the world that they might not have known otherwise. More often than not, one often attempts to infer the external event that caused one's inner states, either verbally or non-verbally. This knowledge of external events in the world might subsequently guide their actions. For instance, in "social referencing" studies, infants used their mothers' expressions to gauge whether to approach or avoid a certain situation. In Sorce et al.'s (1985) study, infants referred to their mothers' expressions to gauge if it was safe or dangerous to cross a visual cliff. If the mother expressed joy or interest, most infants would cross the deep end of the

visual cliff. However, if the mother posed a sad expression, fewer infants crossed. Therefore, infants are capable of identifying the expressions and retrodicting to the relevant circumstance or cause pertaining to the expression which, in turn, regulated their behaviour. As such, if we are able to use other minds as a window onto external events that happened in the world, we can regulate our behaviour accordingly in an adaptive manner.

This concept of retrodiction was previously studied under a related concept, known as "reverse appraisal". As perceivers seem to be aware of the types of relationship between inner states and their relevant situations, they can thus, "reverse engineer" or retrodict to reconstruct the relationship between the behaviour and event based on facial expressions of emotion (Frijda, 1986; Weiner, 2006). From emotion displays, people can retrieve information on how others are appraising an ongoing interaction (the inner state) which then leads to inferences about others' intentions. In De Melo et al. (2014)'s study, participants were told of a certain outcome of the prisoner's dilemma. The video of a target displaying a certain facial expression was shown to perceivers. Perceivers then indicated how they perceived that the target appraised the interaction. From these expressions, perceivers could infer whether the outcome goal was conducive or obstructive to the target. When perceivers saw a smile, it meant that the outcome goal was conducive. But, when participants saw sad, angry or regretful expressions, they inferred that the outcome goal was obstructive for targets. Thus, in this study, perceivers retrodicted from targets' behaviour (i.e. smile) to an external event in the world (i.e. conducive outcome goal), via the inference of targets' inner states (i.e. happy). However, the targets in this study were not real human targets, rather, they were artificial, computer-generated facial expressions. Similar to the

limitations inherent in false belief tasks, there is no prospect of fulfilling the truth condition because there were no real humans or real targets. This study also assumes that every target would react in a similar way to a particular outcome goal. While this study found that people are able to retrodict from facial expressions to an external event that happened to them, the stimuli used were artificial which gives rise to the problem of an objective answer.

More recent studies, however, have investigated people's ability to retrodict based on a novel paradigm which solves the limitations of past studies. A typical paradigm consists of two phases, involving targets and perceivers. Targets' expressions are surreptitiously recorded while they react to a situation or external event, in order to obtain a variety of subtle, dynamic and natural expressions. Then, their expressions are shown to perceivers who are asked to guess the situation or event that targets were reacting to. By asking perceivers to guess the situation or event that targets were reacting to, this task solves the problem of an objective answer because the researcher knows for certain as to which event targets were reacting to. For example, a possible event that might give rise to a smile would be a compliment (Pillai et al., 2012). Upon receiving a compliment, it is assumed that the target is feeling happy. This feeling of happiness is then embodied in the target's expression. In this case, the expression has a proximal cause (i.e., happy) and a distal cause (i.e. compliment). From this display of a happy expression, the perceiver is then able to infer the relevant event that caused the target's expression. In Pillai et al.'s (2012, 2014) study, target participants were recorded surreptitiously when they were being told a joke, compliment, sad story or kept waiting. Their expressions were then shown to perceivers who had to guess what happened to targets. Perceivers were systematically accurate in their responses in

guessing what happened to targets. In another study by Cassidy et al. (2014), targets' expressions were surreptitiously filmed while receiving a chocolate, handmade gift or monopoly money as a reward for carrying out a simple task. When perceivers were asked to guess what gifts targets received, they were systematically accurate in their responses. These studies thus demonstrate that perceivers are systematically accurate at retrodicting from targets' behaviour to an event that happened in the world, presumably by inferring targets' inner states in an intermediate step.

3.1 ADVANTAGES OF RETRODICTION

This novel paradigm for investigating "retrodiction" (Pillai et al., 2012, 2014; Cassidy et al., 2014) enables us to tap into perceivers' ability to make inferences of external events preceding inner states. In other words, it enables us to investigate people's ability to use other minds as a lens onto aspects of the world which cannot be perceived directly.

Unlike past research which used artificial stimuli (Baron-Cohen et al., 1996, 1997; Bartsch and Wellman, 1989; Ekman et al., 1969, 1972; Robinson and Mitchell, 1995; Wimmer and Perner, 1983), this method of retrodiction uses naturalistic and spontaneous behaviour. Because targets are recorded unknowingly, their reactions are dynamic, subtle, natural and spontaneous, similar to the type of expressions seen every day. Since perceivers are required to make inferences from fleeting and spontaneous cues in behaviour, this form of inference represents the moving moment-to-moment demands of social interaction which involves social agents of different personalities, attitudes and thinking, operating

in different situations. As targets' reactions to events are moderated by their personality and inner states such as emotions, intentions, beliefs and knowledge (Flavell et al., 1993), not all targets react in the same way to a particular event. There is a variety of expressions which differs from one another even when targets are experiencing the same inner state. Since there is a high variability in targets' behaviour to one situation, those who are accurate at retrodicting are those who are sensitive to a range of expressions provoked by a single event.

The use of spontaneous and highly varied stimuli means that this paradigm takes into consideration individual differences in how targets behave. In real-life social situations, perceivers would usually consider the personality of targets before attributing an inner state. Even if there are many people reacting in different ways, a perceiver might attribute the same inner state or antecedent event to the range of expressions observed due to the perceiver's understanding of individual differences between targets. Since different targets react in different ways to the same scenario, interpreting targets' behaviour is not trivial because there is no one-to-one correspondence between a specific facial configuration and a situation. For example, when receiving a compliment, some might display a happy expression while others may appear confused, puzzled or embarrassed. Thus, one's reaction to a situation depends, to a certain extent on one's unique set of traits. In short, perceivers' ability to retrodict to an antecedent event (Pillai et al., 2012, 2014; Cassidy et al., 2015) suggests that perceivers understand that there are individual differences in how targets react to a given situation.

Most importantly, this method of retrodiction satisfies the truth condition. In this field of research which focuses on people's ability to infer inner states,

satisfying the truth condition is challenging because we can never really be sure of the true inner state of targets. However, this method bypasses any explicit inference of inner states and rather, allows us to test perceivers' inferences against an objective answer (i.e. the antecedent event) set by the researcher. Since the researcher knows for certain which situations targets were reacting to, perceivers' responses can be measured against an objective answer. If perceivers made an accurate inference of the situation that targets were reacting to, then it can be assumed that they made an accurate inference of targets' inner state. This follows in so far as perceivers infer the proximal cause (i.e. inner states) of the targets' behaviour prior to inferring the distal cause (i.e. situational antecedent) of the behaviour. Thus, if perceivers made an inaccurate inference of targets' inner states, then this would be reflected in their wrong judgements of which situations targets were reacting to. As such, this method of retrodiction satisfies the truth condition articulated by West and Kenny (2011).

Lastly, this method of investigating 'retrodiction' also allows us to disentangle perceiver accuracy from any response bias in perceivers' judgements. Since the spontaneous expressions obtained are often subtle, it is very likely that perceivers would label these spontaneous facial expressions as negative. This bias towards labelling the expressions as negative, however, does not necessarily mean that people are not sensitive in detecting different types of facial expressions. When perceivers are required to make a forced choice from a selection of responses, any bias in their responses might not be due to their ability in making the necessary inferences. Rather, this bias could be caused by a nuisance variable which masks their true underlying ability. Thus, it is important to identify and isolate bias in responses to gain a more accurate impression of people's ability to retrodict to an

external event in the world. A useful procedure for isolating bias is signal detection or Wagner (1993)'s unbiased hit rate. These measures essentially compare hit rates with false alarm rates. For example, *hits* refer to perceivers' correct judgments that targets were responding to a compliment when they really were responding to a compliment while *false alarms* refer to perceivers' judgments that targets were responding to a compliment when targets were, in fact, responding to a joke. Since accounting for false alarms adjusts for response bias, this procedure gives a more accurate and sensitive measure of perceiver's performance.

In summary, this method for investigating 'retrodiction' produces a variety of spontaneous, subtle and naturalistic expressions that we would see in everyday social situations, it accounts for individual differences in people's behaviour, it solves the problem of an objective answer (the truth condition) and it disentangles accuracy from bias. Hence, the findings obtained from 'retrodiction' studies not only paint a more accurate picture of people's ability to make inferences about other minds in everyday situations, but it also enables us to explore people's ability to use other minds as a lens onto external events that happened in the world.

3.2 LIMITATIONS AND CHALLENGES

Although this method of 'retrodiction' has many advantages over previous methods, it is based on the assumption that people infer targets' inner states (in this case, targets' emotions) as a preliminary step to inferring the external event that happened in the world. When perceivers infer an antecedent event from targets' behaviour, it is assumed that perceivers would infer targets' inner states before inferring the antecedent event. However, as past 'retrodiction' studies (Pillai

et al., 2012; Cassidy et al., 2014) did not require perceivers to explicitly reason about targets' inner states, it is not certain if perceivers truly inferred the inner states of targets. Although Cassidy et al. (2014) found that perceivers' emotion inferences corresponded with their inferences of which gifts targets received, these questions relating to emotion and gifts were asked at different points in time and so the evidence does not provide much insight as to whether perceivers truly inferred targets' inner states before they inferred the antecedent event. Since perceivers were not asked to explicitly reason about targets' minds or inner states when retrodicting from targets' behaviour to the event that caused the behaviour, a case can be made for a direct and simple route of processing (Figure 3.2).

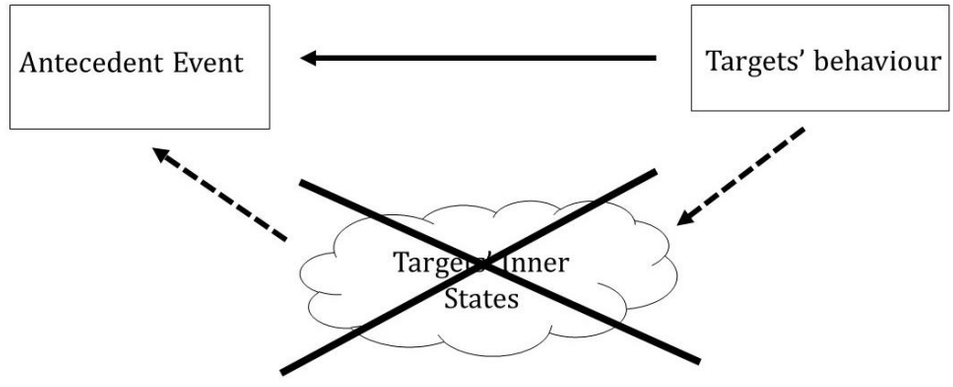


Figure 3.2. Direct and simple route of processing where perceivers retrodict directly from targets' behaviour to the antecedent event, bypassing any inference of targets' inner states.

This proposition for a direct and simple route of processing comes from the longstanding debate between the behaviourist and mentalist explanation for performance on belief tasks. The behaviourist framework states that there is no

need to infer the inner states of others. According to this view, all behaviours are the result of a combination of many different stimuli, organisms and response variables. People's behaviour, in itself, cannot inform other people about the processes or mechanisms that underlie the mental events. Due to the many obscuring and transforming factors between one's behaviour and their mental processes, there is a barrier between both domains (Uttal, 1999). People can choose to hide their true inner state by simulating a different behaviour. In certain cases, people might not even be in touch with their own inner states. Consequently, this would result in a barrier between the targets' true inner state and their behaviour, preventing perceivers from inferring targets' inner states.

In a study of the chimpanzee's acquisition of a 'Theory of Mind', Povinelli and Vonk (2003, 2004) suggested that the chimpanzees did not establish an understanding of the other chimpanzee's mind mediating between looking behaviour and their action of going for the bait. Rather, it can be understood on the basis of how the dominant chimpanzee's visual access influences his likely behaviour, without referring to an inner state of knowledge mediating his visual access and behaviour. The notion that humans predict behaviour by attributing inner states could be exaggerated (Andrews, 2005) as most instances of prediction do not involve mentalistic reasoning (Povinelli and Vonk, 2003). Adults do not typically attribute inner states to predict behaviour. Instead, they can "match" the behaviour to a situation according to a system of behavioural rules (Perner and Ruffman, 2005).

Although this debate originated from research on the representational 'theory of mind', the same criticisms can be applied to people's ability to retrodict.

Similar to the behaviourist view, Gallagher (2008)'s direct perception theory can be used to explain the simple, direct route of processing. Even though this view is specifically about attributing inner states, the same criticism can be applied to 'retrodiction' judgements. For instance, upon viewing a smile, perceivers might infer that the smile is a reaction to a compliment without the need to infer that the target is feeling happy. In this case, perceivers are merely matching targets' behaviour with their corresponding situation according to a system of behavioural rules (Perner and Ruffman, 2005).

Even if targets' responses are varied and there is no simple one-to-one mapping between an expression and a casual event in the world, we can never rule out the possibility that perceivers make retrodictive inferences by associating behaviour with events that occur in the world through a process called 'submentalizing' (Heyes, 2014). If perceivers are not interpreting the expression as an embodiment of their inner states, they are merely classifying the expression of the target. But, inferring what caused an expression uniquely involves retrodictive inference whereas judging if the target's expression is happy or sad does not. In its broadest sense, inferring what caused a behaviour involves a form of mental operation while classifying does not. If retrodiction perceivers matched their responses according to a system of behavioural rules or mentally enacted the event without making any inferences of the target's inner states, then their performance would be similar to perceivers who are merely asked explicitly to classify targets' behaviour.

Another challenge within this field of research is that the conclusions are based on the assumption that people always express what they truly feel. If

perceivers can make an accurate retrodiction judgement, it is assumed that the perceiver has the capacity to use others' minds as a window onto external events that happened in the world. Although the stimuli obtained in this paradigm are spontaneous and genuine, humans are adept at deception as it is a fundamental part of human social interaction (DePaulo et al., 2003). Since our daily social exchange depends on the ability to predict and infer the inner state of others, it is important to judge whether a person is being honest or deceitful. Early studies on deception have focused mostly on verbal deception. When participants were asked to judge whether others are lying or telling the truth from audiotapes and videos, the average accuracy for college students was rarely above 60% (DePaulo et al., 1985; Zuckerman et al., 1981). Custom officials performed similarly to college students in detecting deceit in mock customs examinations (Kraut and Poe, 1980). There was also no difference between federal law enforcement officers, police officers and college students in their ability to discriminate lies from truth (DePaulo and Pfeifer, 1986; Köhnken, 1987). In another study by Ekman and O'Sullivan (1991), only those in the Secret Service performed above chance. As such, people are only marginally accurate at detecting verbal lies.

Verbal lies only form a part of everyday deception. Another common form of deception would be emotional deception. It is possible to infer deception from nonverbal behaviour (Bond et al., 1992; Frank and Ekman, 1997; Vrij et al., 2000) and the best at detecting deceit are those who attended more to nonverbal cues than verbal cues (Ekman and O'Sullivan, 1991). Since aspects of facial expressions are uncontrollable, it can betray a deceiver's true emotion (Ekman & O'Sullivan, 2006). But, emotional deception is rarely studied. People's ability to mask their emotions and simulate a different expression should be taken into consideration to

investigate if people can truly make accurate retrodictive judgements from all types of expressions, including deceptive ones.

In summary, this task of whether people are able to make retrodictive inferences extends past research by providing us with an insight as to whether people can use others' inner states as a window to external events that happened in the world, while solving the limitations of past studies. Nevertheless, we should be cautious in drawing any conclusions that people are indeed inferring others' minds and to take into consideration the fact that people can mask and simulate an expression other than what they are truly feeling.

3.4 AIMS OF THE THESIS

This thesis broadly aims to investigate if people can infer external events that happened in the world from others' behaviour. The first aim of the thesis is to replicate previous findings (Pillai et al., 2012; Cassidy et al., 2014; Teoh et al., 2017) that people can retrodict from others' spontaneous behaviour to a situational antecedent. But, the situational antecedent in this study will be the facial expression of a third party that the targets were looking at. We therefore aim to investigate if perceivers can retrodict from targets' behaviour to the expression of the third party that targets were looking at.

Due to people's ability to deceive others emotionally, the second aim of this thesis is to investigate if people can see through others' simulations to use their minds accurately to infer an external event. In this case, the external event is a picture that targets were looking at. Therefore, we aim to investigate if perceivers

can retrodict from targets' simulated behaviour to the type of picture that targets were really looking at (rather than what expression they were instructed to display).

It is assumed that people make these retrodictive judgements by inferring others' inner states before inferring the event by engaging in an indirect and elaborated processing route. However, as perceivers are not explicitly asked to reason and think about targets' inner states, it is not certain as to whether they are engaging in and indirect and elaborated processing route, or whether they are really engaging in a direct and simple processing route. As such, performance of perceivers who are asked to make 'retrodictive judgements will be compared against that of perceivers who are asked to merely classify targets' behaviour. Retrodicting perceivers are asked to make a response that requires a backwards inference from targets' behaviour to a preceding event that caused the behaviour. On the other hand, classifying perceivers are required to merely classify targets' behaviour, as in a typical emotion recognition task. At the very least, retrodiction involves a form of mental operation while the latter does not involve the form of mental operation required for retrodiction tasks. If retrodiction perceivers engaged in a simple and direct route of processing which does not involve any form of mental operation such as inner state inferences, they would perform similarly as classifying perceivers who were not explicitly asked to retrodict. It is expected that retrodicting perceivers will engage in an indirect and elaborated route of processing while classifying perceivers will engage in a direct and simple route of processing. Thus, the third aim of this thesis is to investigate whether people engage in a simple and direct processing route or elaborated and indirect processing route in making retrodictive judgements.

The three research questions of this thesis are: 1) Can people retrodict from others' spontaneous expressions to an antecedent event (the expression of a person that targets were looking at)? 2) Since people are only marginally accurate in detecting verbal lies and emotional deception, can people retrodict from others' simulated expressions to the true antecedent event (the type of picture that targets were truly viewing)? 3) While retrodicting from others' behaviour to an antecedent event, do people engage in a simple/direct route of processing or an elaborated/indirect route of processing?

CHAPTER 4

Can individuals guess what others were looking at based on brief samples of their behaviour?

EXPERIMENT 1

4.1 INTRODUCTION

The ability to interpret the behaviour of others is a fundamental aspect of social engagement. We often infer the inner states of others in order to understand the social world. Although some information relating to what others think, feel, want or intend comes from verbal communication, we often need to make an inference about others' inner states based on their facial expressions alone. In a classic emotion recognition study by Ekman (1971, 1972), six basic emotions (happy, surprised, sad, angry, disgusted and fear) were shown to be universally recognised from facial expressions. Two decades later, Baron-Cohen et al. (1996, 1997) showed that both typically developing adults and children were able to recognise other complex emotions which extend beyond the six basic emotions. Thus, it is suggested that facial expressions are a visible signal of our inner states and internal experiences, leading to the proposition that we can infer something about another person's mind by interpreting facial expressions.

However, the degree to which past research captures the essence of real life performance is questionable. The stimuli are static, exaggerated and posed. As such, they do not clearly represent the facial expressions seen in daily life. But, the most crucial limitation is that there is no objective answer as to what the actor is actually feeling. For instance, a posed happy expression does not necessarily mean that the actor is feeling happy. According to West and Kenny (2011), one of the most challenging problems in this field of research is knowing the actual inner state of the target such that it can be used as a point of reference in evaluating the accuracy of perceivers' estimation of the target's inner state.

Recent studies have started to overcome these limitations by investigating "retrodiction" from brief and dynamic facial expressions. In making retrodictive inferences (Gallese and Goldman, 1998), the perceiver works backwards from observing the behaviour of a target to inferring the cause of the behaviour, in which perceivers first infer the inner state of targets (proximal cause) before inferring what caused the behaviour (distal cause). This method was developed by Pillai et al. (2012, 2014) and the procedure involves two stages. Firstly, the target is surreptitiously recorded while reacting to a certain situation or stimulus. After that, their videos are shown to perceivers whose task is to infer what caused the reaction of targets.

In Pillai et al. (2012)'s study, perceivers were shown videos of targets reacting to a story, compliment, joke or kept waiting. Perceivers were systematically accurate in guessing what happened to targets. In another study conducted by Cassidy et al. (2014) which employed the same methodology, perceivers were systematically accurate in guessing whether targets received a

handmade gift, chocolate or monopoly money. This paradigm therefore satisfies the truth condition in the sense that the researchers know with certainty what the target was reacting to. Hence, perceivers' judgements can be evaluated against an objective fact.

Furthermore, past studies have focused on the attribution of inner states but this attribution does not fully reflect the broader range of processes and abilities that collectively qualify as the ability to think and infer another person's mind. Being able to think about another person's mind consists of both attribution of inner states and making an inference about inner states (Baron-Cohen et al., 1997).

Moreover, if we can infer something about the world via others' inner states (which we might infer from facial expressions), we can broaden our view of the world. To illustrate, a study by Sorce et al. (1985) showed that infants referenced their mothers' expressions before crossing a visual cliff. Infants only crossed the deep side of the visual cliff if the mother expressed joy or interest. However, if the mother expressed sadness, fewer infants crossed the deep side of the visual cliff. Thus, from one's facial expressions, we are able to know something about the world that we do not already know – in this case, whether an unfamiliar aspect of the environment should be approached or avoided. From Pillai et al. (2012) and Cassidy et al. (2014)'s studies, we know that people can infer the situational antecedent of a particular inner state (the distal cause), apart from attributing the inner state itself (the proximal cause). As such, this method of examining people's ability to retrodict from a facial expression not only informs us about others' inner states, it also greatly extends our knowledge of aspects of the world. Effectively,

people parasitically use other minds as a window to events that happened externally in the world.

In order to explore this possibility of using other minds as a window onto external events, this study investigates perceivers' retrodictive ability in asking perceivers to guess the inner states of a third party. Targets viewed photographs of emotional expressions while being surreptitiously filmed, in case they spontaneously imitated the expressions that they were viewing (Blairy et al., 1999). This spontaneous imitation of expressions has been suggested to reflect a process of emotional contagion (Laird et al., 1994), whereby the observation of emotional expressions will first trigger the corresponding emotion in the observer which subsequently elicits the same facial expression (Lundqvist and Dimberg, 1995). This process of imitation would consequently lead to an accurate emotion recognition by allowing the participant to access stored knowledge about the emotional states associated with the facial expressions (Wood et al., 2016).

While viewing the expressions, targets were asked to either memorise the expressions or imagine how the actor was feeling. According to the emotional contagion hypothesis, targets will imitate the expressions that targets were viewing, regardless of whether they were asked to memorise the expressions or imagine how the actor was feeling. Based on these recordings of targets' reaction to emotional expressions, perceivers were then asked to infer what category of facial expressions targets were looking at. Similar to Pillai et al.'s (2012) and Cassidy et al.'s (2014) studies, it is expected that targets will retrodict from targets' behaviour to what category of expressions they were looking at, whereby the

proximal cause is the inner state of targets and the distal cause is the expression that they were looking at.

4.2 METHODS

The study was divided into two parts; Generation of Target Video Clips and Perceiver Phase. The study was approved by the Ethics Committee, School of Psychology, University of Nottingham.

4.2.1 Generation of Target Video Clips

The aim of this part of the study was to generate dynamic, subtle and brief reactions. As the representation of other inner states is necessary in order to recognise and comprehend social emotions, targets were shown emotional expressions in order to elicit a reaction which could embody forms of social emotion. As a cover, they were told that they were participating in a memory test. In one condition, they were asked to memorise the expression while in another condition, they were asked to imagine what the person was feeling. These reactions were filmed and perceivers were asked to determine which category of facial expression targets looked at. As such, this task satisfies the truth condition as the researcher is aware of the precise category of emotional expressions that targets were looking at.

4.2.1.1 Target participants

Forty participants (20 females, 20 males, Mean age= 20.03 years, SD= 2.06 years) from the University of Nottingham Malaysia Campus were recruited as target participants. They were recruited from the undergraduate and postgraduate

student population. The targets were given chocolate as a form of inconvenience allowance.

4.2.1.2 Materials and Apparatus

Targets were filmed in a spacious experimental room within the School of Psychology. They were seated with their backs facing the door and against the wall. As soon as the recording began, the researcher left the experimental room.

Targets' expressions were surreptitiously filmed using the webcam of a Toshiba Satellite CS650-S5501 laptop. The laptop was positioned approximately 60 centimetres (cm) from the target's seat. The laptop was also positioned in such a way that the laptop webcam recorded the target's face, neck and shoulders against a white background.

The emotional expressions which targets had to view were obtained from the "Pain" set of faces from the Psychological Image Collection at Stirling (PICS) faces database (https://pics.stir.ac.uk). This set consists of 599 posed, static and exaggerated images, of which 13 were females and 10 were males. Each actor had to display at least two expressions of each of the six basic emotions. Only 5 female and 5 male expressions of the six basic emotions were selected as stimuli by randomization in Microsoft Excel 2007, amounting to a total of 60 emotional expressions. The experiment was presented on PsychoPy version 1.8 (Peirce, 2007).

4.2.1.3 Experimental and Filming Procedure

Targets were told that they would be participating in a recognition memory test for faces. It was a between-groups design where one group of targets were

asked to memorise the expression they viewed ("memory") while another group of targets were asked to imagine what the actor was feeling ("imagine"). The laptop webcam was set to record targets' expressions as soon as targets were seated. In order to avoid any suspicion that they were being recorded; the webcam light was covered with opaque stickers which supposedly formed part of the decoration on the laptop. This was to avoid any unnecessary attention to the webcam light. Each emotional expression was displayed for 5 seconds. Targets had to view 120 posed and static emotional expressions in total, whereby each condition consisted of 60 emotional expressions. Both conditions consisted of the same set of emotional expressions in order to avoid any potential confounding factors such as the intensity of expressions. After the experiment, targets were debriefed about the actual purpose of the experiment. All targets completed a video release consent form after the experiment. If target participants agreed for their videos to be used, they would have to complete the form. If, however, they do not agree for their videos to be used, their videos and personal information would be destroyed immediately.

4.2.2 Editing and Selecting Video Clips

Targets' expressions were edited using Windows Movie Maker 2.6. Targets' videos were segmented into a series of 3-seconds video clips which corresponded with the expression that they were looking at and to capture any target reactions. The video frames were 400 pixels in width and 300 pixels in height. The audio component of the videos were omitted. There were 20 sets of video clips in the "memory" condition and 20 sets of video clips in the "imagine" condition. As there were 40 sets of video clips and 60 reactions in each set, a library of 2400 individual video clips was created. These individual video clips were randomised in Microsoft

Excel 2007 and only 120 video clips of targets' expressions were selected. Half of these video clips depicted targets memorising the expressions that they were looking at and another half of the video clips depicted targets imagining the expressions that they were viewing. There were an equal number of female and male targets. As these video clips were randomized in Microsoft Excel, it should be noted that there was an unequal amount of reactions displayed by each target. Thus, the video clips in the "memory" condition consisted of 17 different targets (average of 3.53 reactions per target) and the video clips in the "imagine" condition consisted of 15 different targets (average of 4 reactions per target). However, there were no repetitions of targets for each type of reaction.

[Figure 4.1 was removed to maintain confidentiality]

Figure 4.1. Sample screenshot of target reactions in the "memory condition" to (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) fear, (f) sad expressions.

[Figure 4.2 was removed to maintain confidentiality]

Figure 4.2. Sample screenshot of target reactions in the "imagine condition" to (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) fear, (f) sad expressions.

4.2.3 Perceiver Phase

4.2.3.1 Participants

Twenty participants (10 females, 10 males, *Mean age*=19.80 years, *SD*=4.20 years) aged between 18 to 30 years old took part in this study. The experiment was conducted at the University of Nottingham Malaysia Campus. All participants were given chocolates as a form of inconvenience allowance and written consent was obtained.

4.2.3.2 Materials and Apparatus

One-hundred-and-twenty video clips of targets' subtle, dynamic and natural reactions to emotional expressions were used as stimuli to present to perceivers. A series of generic emoticons were obtained from the internet to represent each category of the six basic emotions. These emoticons were used as response options

to represent each category of the six basic emotional expressions targets were looking at. Emoticons were used in place of Ekman expressions from the KDEF database to avoid the possibility of perceivers looking for a reaction to the particular Ekman expression. Verbal labels were not used in order to minimise overcrowding of response options on the screen.

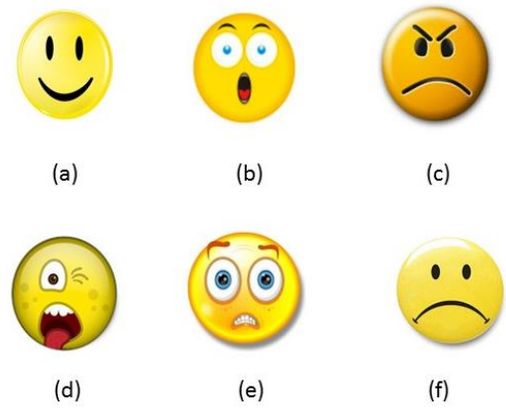


Figure 4.3. Emoticons used to represent the six basic emotions: (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) fear and (f) sad.

The experiment was conducted on PsychoPy v 4.80 (Peirce, 2007) and displayed on a Toshiba Satellite C655-S5501 laptop. The stimuli were randomized using PsychoPy. The order of the response options always remained constant (from left to right: happy, surprised, angry, disgusted, fear, sad). Although the duration of the stimuli was only three seconds, the response options were displayed until perceivers selected their response.

4.2.3.3 Design

This experiment was a within-subjects design in that all perceivers viewed targets' reactions to emotional expressions in the "memory" and "imagine" condition.

4.2.3.4 Procedure

Perceivers were seated approximately 60 cm from the laptop screen. Before the experiment began, perceivers were told that they will be viewing reactions to pictures of emotional expressions. Perceivers first started with an example trial in which they saw an example of a target's expression. They were asked to select the category of facial expression that they thought targets were looking at. Perceivers were asked to indicate their response by clicking on the emoticon that they thought best represented the category of facial expression that targets were viewing. They were also told the emotional categories which corresponded with the emoticons. Perceivers were told to respond as quickly and accurately as possible. After the example trial, they started with the actual testing phase which was similar to the example trial. The target expression which was used in the example trial was not used in the test trial. All video clips were randomized. After the experiment, all perceivers were debriefed.

4.3 RESULTS

The results were analysed using the Statistical Package for Social Sciences (SPSS) 16.0 software.

4.3.1 Were perceivers systematically accurate in guessing what category of facial expressions targets were looking at?

In order to determine if there was any bias in perceivers' responses, a confusion matrix was created, as shown in Table 4.1

Table 4.1 *Confusion matrix of the percentage of correct and incorrect inferences (and data count) in both conditions. Shaded cells represent the correct responses.*

		Stimulus						
		Happy	Surprised	Angry	Disgusted	Fear	Sad	Total responses
Memory condition	Happy	57.50 (115)	18.00 (36)	11.50 (23)	13.50 (27)	11.00 (22)	9.00 (18)	120.50 (241)
	Surprised	9.00 (18)	21.00 (42)	14.50 (29)	13.50 (27)	17.00 (34)	7.50 (15)	82.50 (165)
	Angry	7.50 (15)	12.50 (25)	20.50 (41)	15.50 (31)	13.00 (26)	24.00 (48)	93.00 (186)
	Disgusted	8.00 (16)	4.50 (9)	11.50 (23)	14.50 (29)	11.50 (23)	3.00 (6)	53.00 (106)
	Fear	6.00 (12)	21.50 (43)	16.50 (33)	15.50 (31)	20.50 (41)	12.00 (24)	92.00 (184)
	Sad	12.00 (24)	22.50 (45)	25.50 (51)	27.50 (55)	27.00 (54)	44.50 (89)	159.00 (318)
	Total	100.00 (200)	100.00 (200)	100.00 (200)	100.00 (200)	100.00 (200)	100.00 (200)	
Imagine condition	Happy	49.00 (98)	15.50 (31)	9.00 (18)	21.50 (43)	8.50 (17)	10.00 (20)	113.50 (227)
	Surprised	10.50 (21)	21.00 (42)	10.50 (21)	13.00 (26)	12.50 (25)	16.50 (33)	84.00 (168)
	Angry	7.50 (15)	6.50 (13)	16.50 (33)	12.50 (25)	16.50 (33)	11.50 (23)	71.00 (142)
	Disgusted	4.50 (9)	10.50 (21)	10.00 (20)	22.00 (44)	8.50 (17)	6.50 (13)	62.00 (124)

	Fear	10.00 (20)	17.00 (34)	19.50 (39)	17.00 (34)	15.50 (31)	18.00 (36)	97.00 (194)
	Sad	18.50 (37)	29.50 (59)	34.50 (69)	14.00 (28)	38.50 (77)	37.50 (75)	172.50 (345)
	Total	100.00 (200)	100.00 (200)	100.00 (200)	100.00 (200)	100.00 (200)	100.00 (200)	

As there was a bias towards responding that a given expression is a reaction to a sad emotional expression, unbiased hit rates (H_u) (Wagner, 1993) were calculated to determine if perceivers could discriminate between reactions to the emotional expressions. This method expresses accuracy as proportions of both response frequency and stimulus frequency. As such, it is insensitive to bias, the different proportions of different types of stimuli and the different number of categories.

Table 4.2 *Example confusion matrix based on Wagner (1993)'s scheme.*

Responses	Stimulus			
	1	2	3	Total
1	A	D	g	a + d +g
2	B	E	h	b + e + h
3	C	F	i	c + f + i
Total	a + b + c	d + e + f	g + h + i	N

According to the example confusion matrix shown in Table 4.2, the formula for calculating H_u for stimulus category *a* is as follows:

$$H_u = \frac{\mathrm{a}}{\mathrm{a}+\mathrm{b}+\mathrm{c}} \times \frac{\mathrm{a}}{\mathrm{a}+\mathrm{d}+\mathrm{g}}$$

When the stimulus type is correctly identified by all of the responses, H_u has a value of +1. If all of the responses do not correspond with the stimulus category, it is simply entered as 0. This method of calculation has been used extensively in previous research involving emotion recognition (for example, Elfenbein and Ambady, 2003; Orgeta et al., 2010; Pell et al., 2009; Tcherkassof et al., 2007).

The probability of selecting a particular category by chance (P_c) was calculated for each category of stimulus using the following formula (based on Table 4.2).

$$P_c = \frac{\mathrm{a}+\mathrm{b}+\mathrm{c}}{\mathrm{N}} \times \frac{\mathrm{a}+\mathrm{d}+\mathrm{g}}{\mathrm{N}}$$

Mean H_u was calculated by averaging each perceivers' individual H_u scores and Table 4.3 shows the Mean H_u and respective P_c for each category of targets' reactions.

Table 4.3 *Mean (and standard deviation) of unbiased hit rates (H_u) and chance probability (P_c) to targets' reactions to each emotional expression.*

Expression Viewed	Target's condition			
	"Memory" condition		"Imagine" condition	
	Mean H_u	P_c	Mean H_u	P_c
Happy	0.30 (0.12)	0.03	0.22 (0.08)	0.03
Surprised	0.08 (0.07)	0.02	0.07 (0.06)	0.02
Angry	0.05 (0.05)	0.03	0.05 (0.05)	0.02
Disgusted	0.05 (0.05)	0.01	0.10 (0.08)	0.01
Fear	0.06 (0.08)	0.03	0.04 (0.04)	0.03
Sad	0.14 (0.09)	0.04	0.10 (0.06)	0.05

One-sample *t*-tests were conducted in order to determine if perceivers' judgements were significantly above chance probability. Regardless of what targets were asked to do, H_u was significantly above chance probability for reactions to happy [$t(19)=12.33$, $p<0.001$], surprised [$t(19)=4.24$, $p<0.001$], sad [$t(19)=3.95$, $p=0.001$], angry [$t(19)=12.33$, $p<0.001$] and disgusted [($t(19)=5.24$, $p<0.001$] expressions, but not fear [$t(19)=1.80$, $p=0.087$]. One-sample t-tests were also conducted separately, according to targets' conditions. When targets were memorising the emotional expressions that they viewed, perceivers were significantly above chance in guessing their reactions to happy [$t(19)=9.65$, $p<0.001$], surprised [$t(19)=3.45$, $p=0.003$], angry [$t(19)=2.37$, $p=0.028$], disgusted [$t(19)=3.05$, $p=0.007$], fear [$t(19)=4.92$, $p<0.001$] and sad [$t(19)=3.95$, $p<0.001$] expressions. When targets were asked to imagine how the actor was feeling, perceivers were systematically above chance in guessing targets' reactions to happy [$t(19)=10.69$, $p<0.001$], surprised [$t(19)=3.44$, $p=0.003$], angry [$t(19)=2.48$, $p=0.023$], disgusted [$t(19)=4.39$, $p<0.001$] and sad [$t(19)=3.95$, $p=0.001$] but not fear [$t(19)=1.16$, $p=0.26$].

4.3.2 Were perceivers' responses affected by what targets were asked to do?

In order to identify if perceivers' responses were affected by what targets were asked to do, a two-way repeated measures Analysis of Variance (ANOVA) was conducted on the H_u scores with Target Condition (memory, imagine) and Expression Viewed (happy, surprised, angry, disgusted, fear, sad) as within-subjects factors. There was a significant main effect of "Expression Viewed"

[F(5,95)=43.64, p<0.001, η_p^2=0.70]. Pairwise comparisons (with Bonferroni correction) revealed that perceivers were more accurate in judging reactions to happy expressions than reactions to surprised (p<0.001), angry (p<0.001), disgusted (p<0.001), fear (p<0.001) and sad (p<0.001) expressions. Perceivers were also significantly better at judging reactions to sad expressions as compared to angry *(*p=0.009), disgusted (p=0.033) and fear (p=0.001). There was a significant interaction between "Target Condition" and "Expression Viewed", [F(5,95)=3.47, p=0.006, η_p^2=0.16]. The graph depicting this interaction is shown in Figure 4.4. There was no significant main effect of "Target Condition" [F(1,19)=4.01, p=0.06, η_p^2=0.17].

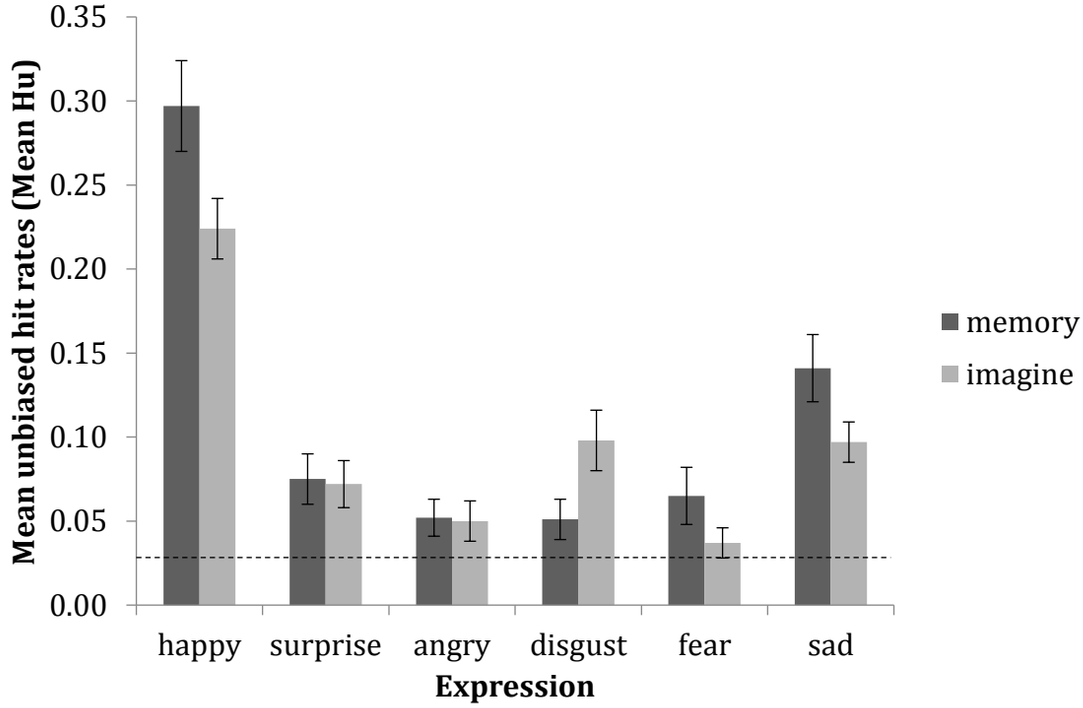


Figure 4.4. Interaction between 'Target Condition" and "Expression Viewed". Error bars represent the standard error of mean. The dotted line represents average chance probability (P_c) at 0.03.

4.3.2.1 Post-hoc analyses for "Target Condition" and "Expression Viewed" interaction

In order to further explain the interaction between "Target Condition" and "Expression Viewed", two one-way ANOVAs were conducted. When targets were asked to memorise the expression that they were viewing, there was a significant main effect of "Expression Viewed", $F(5, 95)=28.17$, $p<0.001$, $\eta_p^2=0.59$. Pairwise comparisons (with Bonferroni correction) revealed that perceivers were systematically more accurate in guessing happy as compared to surprised ($p<0.001$), angry ($p<0.001$), disgusted ($p<0.001$), fear ($p<0.001$) and sad ($p=0.004$). Perceivers were also systematically more accurate in guessing sad as compared to angry ($p=0.044$), disgusted ($p=0.017$) and fear ($p=0.028$). When targets were asked to imagine what the actor was feeling, there was a significant main effect of expression viewed, $F(5, 95)=21.24$, $p<0.001$, $\eta_p^2=0.53$. Pairwise comparisons (with Bonferroni correction) revealed that perceivers were systematically more accurate in guessing happy as compared to surprised ($p<0.001$), angry ($p<0.001$), disgusted ($p<0.003$), fear ($p<0.001$) and sad ($p<0.001$). Perceivers were also systematically more accurate in guessing sad as compared to fear ($p=0.004$). Six paired-samples *t*-tests (with Bonferroni correction) showed that there was no significant difference in perceivers' responses between both target conditions across all types of reactions [Happy, $t(19)=2.64$, $p=0.02$; Surprise, $t(19)=0.19$, $p=0.85$; Sad, $t(19)=0.14$, $p=0.89$; Angry, $t(19)=-2.06$, $p=0.05$; Disgust: $t(19)=1.40$, $p=0.18$ and Neutral, $t(19)=1.76$, $p=0.095$].

4.3.3 Were perceivers' response times affected by the behaviour of targets?

Outliers 2.5 ± times the median were identified and 6 out of 240 (2.50%) individual cases were removed, according to the Median Absolute Deviation (Leys et al., 2013). The assumption of normality was not violated for all variables, *Shapiro-Wilks, p*>0.05.

In order to establish whether perceivers' response times were affected by the expression targets viewed and what they were instructed to do, a two-way repeated-measures ANOVA was conducted on perceivers' reaction times with Target Condition (memory, imagine) and Expression Viewed (happy, surprised, angry, disgusted, fear, sad) as within-subjects factors. The descriptive statistics are shown in Table 4.4.

Table 4.4 *Mean (and standard deviation) for perceivers' response time in miliseconds (ms) for both target conditions.*

Expression viewed	Target Condition	
	"memory"	"imagine"
Happy	5978.38 (1617.77)	5422.56 (1183.97)
Surprised	5140.40 (1230.17)	5220.59 (946.23)
Angry	5853.63 (1747.28)	6065.58 (1393.63)
Disgusted	6029.01 (1226.55)	5981.22 (1156.78)
Fear	5464.86 (1063.28)	5931.91 (1555.88)
Sad	5697.37 (1363.38)	5872.98 (1429.91)

There was a significant main effect of "Expression Viewed" [$F(5, 80)=4.56$, $p=0.001$, $\eta_p^2=0.22$]. Pairwise comparisons with Bonferroni correction revealed that perceivers were significantly faster at judging reactions to surprised expressions than reactions to happy ($p=0.037$), angry ($p=0.003$) and disgusted ($p=0.005$) expressions. But, there was no significant main effect of Target Condition [$F(1, 16)=0.23$, $p=0.64$, $\eta_p^2=0.01$] and no significant "Target Condition" and "Expression Viewed" interaction [$F(5, 80)=1.29$, $p=0.28$, $\eta_p^2=0.08$].

4.4 DISCUSSION

The findings revealed that perceivers could systematically discriminate between reactions to different emotional expressions. Although they were systematically accurate in their responses, they were only slightly above chance. When targets were asked to memorise what they were viewing, perceivers could systematically discriminate between reactions to all emotional expressions. When targets were asked to imagine the expressions that they viewed, perceivers could systematically discriminate between reactions to most emotional expressions except fear. Although these results may indicate that perceivers were systematically more accurate when targets were asked to memorise, as compared to when they were asked to imagine what the actor was feeling, there was no significant difference in perceivers' judgements between the two conditions.

Perceivers were significantly above chance in inferring what category of facial expressions targets were looking at, presumably because targets spontaneously imitated the emotional expression they were viewing. This finding

is consistent with past studies which showed that people spontaneously imitate others' expressions (Blairy et al., 1999; Lundqvist and Dimberg, 1995; Dimberg et al., 2000; Sato and Yoshikawa, 2007; Wood et al., 2016). This spontaneous imitation may reflect an internal simulation of perceived emotion which may facilitate the understanding of others' emotions (Niedenthal et al., 2001). The results also indicated that perceivers were systematically more accurate in guessing reactions to a happy expression as compared to all the other emotions, presumably because the imitation of a happy expression is more distinctive than the imitation of other expressions. Because a happy expression can be characterised by only a smiling mouth (Adolphs, 2002), a complete visual analysis of the expression was perhaps not needed (Fabre-Thorpe et al., 2001; Rodger et al., 2015). On the other hand, negative emotions may be more difficult to produce (Öhman et al., 2001), resulting in poorly imitated negative expressions.

As a whole, these findings suggest that perceivers can make retrodictive inferences from brief samples of behaviour, consistent with Pillai et al. (2012) and Cassidy et al. (2014)'s findings. When perceivers viewed targets' reactions, they made a retrodictive inference from the expressions to the emotional expression that targets viewed, whereby the proximal inference is the inner state (in this case, emotion) of the target.

But, some may argue that perceivers were merely classifying targets' reactions, without inferring the underlying inner state. For instance, when perceivers saw a target smiling, he or she would recognize that the expression is a reaction to a happy expression without inferring that the target is actually feeling happy. This is known as Povinelli's challenge (Povinelli and Vonk, 2003).

Furthermore, Gallagher (2008), according to the Direct Perception theory, also suggested that we do not need to think about other people's minds as perception alone provides sufficient information for understanding others. For example, when we encounter someone smiling, we would think of it as a happy expression and not that the person is feeling happy. But, a reaction to an emotional expression does not reflect a perfect or consistent imitation of that particular emotional expression across all targets. Thus, the variety of reactions evoked by a particular emotional expression might make it difficult for perceivers to match it according to a classification template. Furthermore, some reactions may be very subtle and there may not be much observable facial clues to match or classify the expression. Nevertheless, as there was no comparison against a control task in which perceivers were not asked to infer targets' inner states, these alternative explanations cannot be rejected.

In summary, the findings from this study revealed that perceivers were above chance level in inferring the inner state of a third party signalled in targets' expressions, indicated by perceivers' judgments of which category of facial expressions targets were looking at. However, the present findings are insufficient to conclude that perceivers engaged in an elaborated and indirect processing strategy which requires the inference of inner states. Thus, the next study focused on replicating the present results of whether perceivers can infer the signalled inner state of a third party and additionally, the processing strategy involved.

CHAPTER 5

Differences in processing strategies between retrodiction and classification based on spontaneous facial expressions

EXPERIMENT 2:1

5.1 INTRODUCTION

From many decades of research on people's ability to understand inner states, we know that people are exceptionally good at inferring basic (Ekman et al., 1969, 1972) and complex (Baron-Cohen et al., 1996, 1997) emotions from facial expressions. A recurring problem with these studies is that the stimuli used are often posed and exaggerated expressions where there is no objective answer as to what the actor is truly feeling. Moreover, people's ability to infer and recognise others' inner states may be attributed to the exaggerated cues in the actor's expressions, thus reducing any chances of making an error. Since natural facial expressions in everyday life are often subtle, these findings do not necessarily paint a true picture of people's ability to infer others' inner states in everyday social communication. Due to these exaggerated cues, arguably, it might be possible to classify these expressions without having to think about others' inner states (Gallagher, 2008). For instance, upon seeing a smiling face, the observer might think that it is a happy expression rather than thinking that the actor is genuinely

feeling happy. Even if people are genuinely thinking about others' inner states, this ability to infer others' emotional states is just one aspect of the various inferences which could be inferred from one's mind.

Following on from that, inner states do not occur alone but are usually preceded by a certain event that is unobservable to the person observing the inner state. These events do not necessarily have one-to-one correspondence with a certain inner state as a particular event may elicit different inner states across different people. Thus, given that these external events give rise to different inner states which are signalled in others' behaviour, can people infer these unobservable events from a brief sample of others' behaviour? In short, can people use other minds as a window onto external events that happened in the world? This question can be addressed by considering a process known as "retrodiction". In this procedure that investigates the process of retrodiction, targets are surreptitiously filmed while reacting to a situation. Then, their reactions are shown to perceivers who will have to infer what situation targets were reacting to. This method of "retrodiction" solves many problems faced by past research. Firstly, the expressions obtained are subtle and dynamic since targets are recorded surreptitiously. More importantly, it solves the problem of an objective answer because the researchers know for certain which situations targets were reacting to, avoiding any uncertainty as to what the true internal state of targets were. Furthermore, this method also allows us to tap into a different aspect of whether people are able to infer unobservable external events that happened in the world, from one's naturalistic behaviour. To date, "retrodiction" studies have shown that perceivers are exceptionally accurate at guessing whether targets were reacting to a joke, compliment, story, kept waiting (Pillai et al., 2012); whether targets

received chocolate, monopoly money, handmade gift (Cassidy et al., 2014); targets' preferences of attractiveness of people, humour of cartoons, decorative appeal of paintings and cuteness of animals (North et al., 2012); and more recently, whether targets were accompanied or alone (Teoh et al., 2017).

Other than the events mentioned above, another plausible cause of one's inner state is the automatic imitation of another person's inner state (Dimberg et al., 2000; Sato and Yoshikawa, 2007; Wood et al., 2016). If another person's inner state is embedded in their expressions, surely it is possible for people to guess the inner state viewed. For example, in a situation where person A is communicating with person B while person C looks from a distance in such a way that person A's expression is blocked from C's view, C cannot directly infer A's inner state. If A's inner state is reflected in B's expression, C can thus infer A's inner state by looking at B's expression. In order to investigate whether people can infer the inner state that others are looking at, we adapted the 'retrodiction' procedure as such: - Targets were surreptitiously filmed while looking at pictures of emotional expressions. As a cover activity, they were told to choose an emoticon which best represents the expressions viewed. Their reactions were then shown to perceivers who had to guess which emoticons targets chose.

If perceivers made an accurate judgement, it is generally assumed that perceivers first inferred the inner states of targets (proximal cause) which reflected what they were viewing (distal cause). Once perceivers inferred what pictures targets were viewing, they can use that information to guide their judgement of which emoticons targets chose. This explanation follows an elaborated and indirect route of processing, as shown in Figure 5.1. In other words,

by selecting the appropriate emoticon, perceivers are accurate at inferring the picture that was viewed by targets. For example, upon seeing a smiling face, perceivers would infer that targets are feeling happy, thus they must be looking at a happy face and therefore, chose a happy emoticon to represent what they (targets) were viewing. An alternative explanation would be that perceivers retrodict from targets' behaviour directly to the emoticons chosen by targets, bypassing any inference of targets' inner states. If perceivers merely chose the emoticons that targets chose without inferring the inner states of targets, it is similar to choosing the emoticon which best represents targets' expressions, analogous to a simple and direct route of processing (as shown in Figure 5.2).

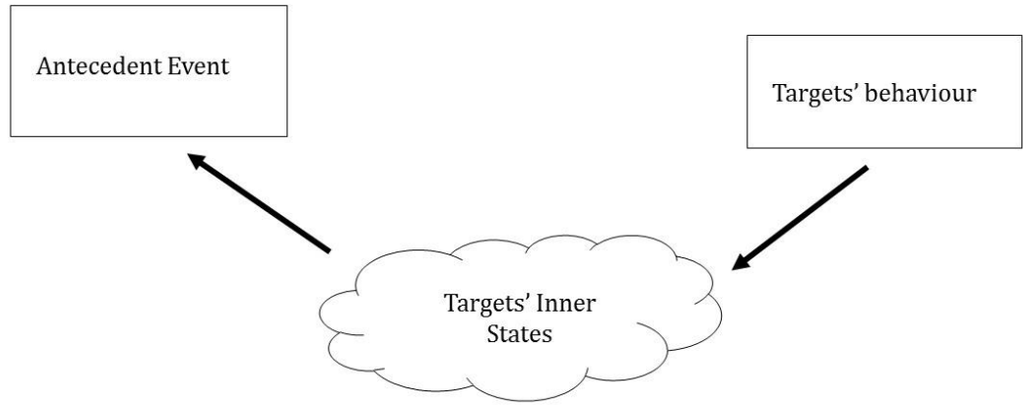


Figure 5.1 Indirect and elaborated route of processing whereby perceivers infer targets' inner states before inferring the antecedent event.

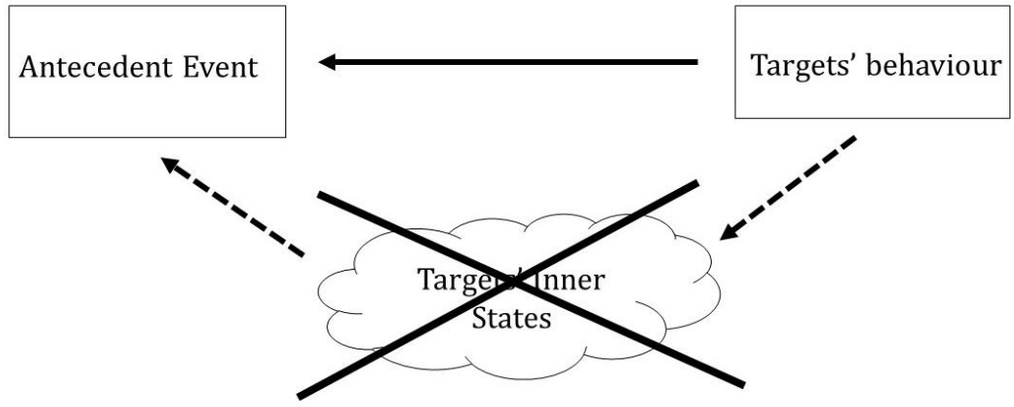


Figure 5.2 Direct and simple route of processing where targets directly infer the situational antecedent from targets' behaviour, without any inference of targets' inner states.

As such, to investigate if perceivers were genuinely retrodicting through an elaborated and indirect route of processing, the performance of retrodicting perceivers will be compared against the performance of classifying perceivers. Retrodicting perceivers were asked to choose the emoticons that targets chose while classifying perceivers were asked to select the emoticon that best represents targets' expressions. The main difference between these two tasks is that the former task requires perceivers to explicitly think about target minds whereas the latter does not require perceivers to explicitly think or consider about targets' minds. The use of natural and subtle expressions might increase the chances of making an error by classifying perceivers since these expressions may not provide sufficient cues for them to accurately classify the expressions, thus, disentangling the processes involved in both groups. If retrodiction perceivers processed their responses in the same way as perceivers in the classification condition, then perceivers' responses in both groups should be the same. We expect that retrodiction perceivers will be systematically accurate at inferring which

emoticons targets chose, reflecting their ability to infer the inner states (in this case, emotional states) that was viewed by targets. In doing so, we expect retrodiction perceivers to be using an elaborated and indirect route of processing while classifying perceivers use a simple and direct route of processing.

5.2 METHODS

The study was divided into two parts; Generation of Target Video Clips and Perceiver Phase. Ethical approval was obtained from the Ethics Committee, School of Psychology, University of Nottingham.

5.2.1 Generation of Target Video Clips

The aim of this part of the study was to generate dynamic, subtle and brief reactions to emotional expressions. Similar to Experimental Chapter 1, targets were shown emotional expressions in order to produce a reaction which reflects the embodiment of social emotion. As a cover activity, targets were asked to select the emoticon which best represents the expression they were looking at. Their reactions to the emotional expressions were surreptitiously filmed. Perceivers in both groups were shown these video clips. In the retrodicting group, perceivers were asked to guess which emoticon targets chose. Meanwhile, perceivers in the classification group were asked to select the emoticon which best represents the target's expression. Responses made by perceivers could then be compared against an objective fact – the expression targets were actually looking at and/or the emoticon actually selected by the target as they were viewing the expression.

5.2.1.1 Target participants

Thirty participants, both undergraduate and postgraduate students at the University of Nottingham (15 females, 15 males, Mean age = 20.27 years, SD = 5.36 years) were recruited as targets. They were given £3 as an inconvenience allowance.

5.2.1.2 Materials and Apparatus

Targets were filmed in a spacious research lab within the School of Psychology. They sat against a white background. The researcher left the lab as soon as the experiment began. Targets' expressions were surreptitiously filmed with a Toshiba Satellite CS650-S5501 laptop. The laptop was positioned approximately 60 cm from the target's seat. It was positioned in such a way that the laptop webcam recorded the target's face, neck and shoulders.

The emotional expressions which targets viewed were obtained from the Karolinska Directed Emotional Faces (KDEF) database (Lundqvist et al., 1998) which consists of a total of 4900 pictures of facial expressions of emotion. This database consisted of 70 actors, 35 of whom were females and 35 of whom were males. Each actor had to pose 7 different expressions (neutral, happy, angry, afraid, disgusted, sad, surprised) at 5 different angles (full left profile, half left profile, straight, half right profile, full right profile). The size of these photographs was 562 x 762 pixels while the resolution was 72 x 72 dpi. Two copies of each type of posed expressions were obtained, amounting to a total of 70 photographs. For the purpose of this experiment, only 5 female and 5 male actors were selected at random using MS Excel 2007. Photographs of six different expressions (happy, surprised, sad, angry, disgusted, neutral) shot at a straight angle were selected for

the target phase. Fear was excluded due to its confusion with surprised (Gosselin and Kirouac, 1995; Jack et al., 2009) as both fear and surprise have visually similar features (Gagnon et al., 2010; Gosselin and Simard, 1999). If targets mistakenly perceived fearful expressions as surprised expressions, they might choose the surprised emoticon to represent the fearful expression. Even if perceivers' response that targets chose a fearful emoticon is accurate, it is incongruent with what targets were actually viewing. In such cases, an accurate response of what targets chose does not reflect a true depiction of what targets were viewing. Hence, fear was excluded to avoid any potential discrepancy between what targets' responses and what they were actually viewing. The emoticons used were obtained from the internet. They represented the emotional category of happy, surprised, sad, angry, disgusted and neutral and were displayed in the same order (from left to right: happy, surprised, sad, angry, disgusted, neutral). The size of the emoticons was 76 x 76 pixels. Figure 5.3. shows the emoticons used to represent each emotional category. The experiment was displayed with PsychoPy v.1.80.06 (Peirce, 2007).

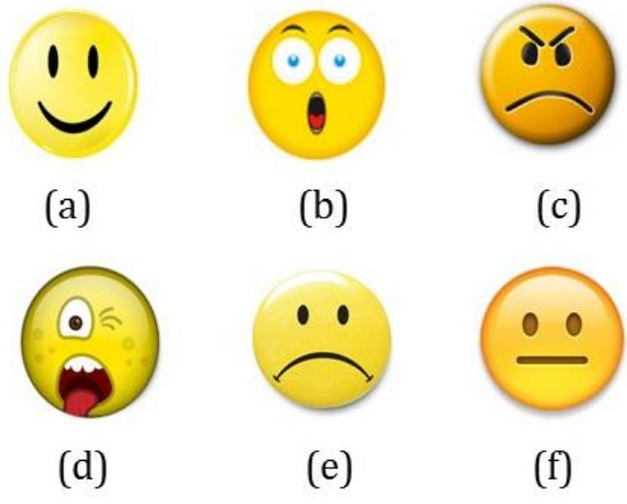


Figure 5.3. Emoticons used to represent the emotional categories: (a) happy, (b) surprised, (c) angry, (d) disgusted, (e) sad and (f) neutral.

5.2.1.3 Experimental Procedure and Filming

Similar to the procedure used in Experiment 1, targets were told that they would be participating in a recognition memory test for faces. They were asked to memorise the expression that they viewed. In addition, they were also asked to select the emoticon which best represents the displayed expression. In order to avoid any suspicion that they were being filmed, the webcam light was concealed with some opaque stickers and a label stating "FOR EXPERIMENTAL PURPOSES ONLY." Each photographed emotional expression was displayed for 5 seconds. Targets viewed 60 static and posed facial expressions from the KDEF database. After the experiment, they were debriefed about the true purpose of the experiment. If targets agreed for their videos to be used, they were asked to complete a video release consent form. However, if they did not agree for their videos to be used, their videos and personal information would be destroyed immediately.

5.2.2 Editing and Selection of Video Clips

According to the data obtained from PsychoPy (Peirce, 2007), target video clips were edited to correspond with the expression that they viewed. As there were 30 targets each viewing 60 emotional expressions, there were 1800 target video clips in total. As targets made their responses of selecting which emoticon represents the expression viewed within 3 seconds, the duration of targets' expressions were trimmed to capture the first three seconds of targets' expressions in order to standardize the duration across all emotional categories. The video clips

were 640 pixels in width and 480 pixels in height. The audio component of the video clips was omitted. From the 1800 individual clips, 77 were selected on the grounds that there was visible facial movement in the target. The videos were rated for facial movement by a second person and the inter-rater agreement was *Kappa* = 0.83 (p<0.001). The remaining 1723 individual clips were entered into MS Excel and 163 videos were selected at random. Therefore, there was a total of 240 video clips which were used in the validation stage.

5.2.2.1 Validation of stimuli

Ten judges (5 females, 5 males) were asked to assess the expressiveness of the video clips from a scale of 1 (least expressive) to 6 (most expressive). Based on the expressiveness ratings, the 10 least expressive and 10 most expressive video clips for each emotion were chosen for the perceiver phase. For each of these video clips, the emotional category that targets viewed corresponded to the emotional category of emoticons targets chose to represent the expression. The selection and validation stage is summarised in Figure 5.4.

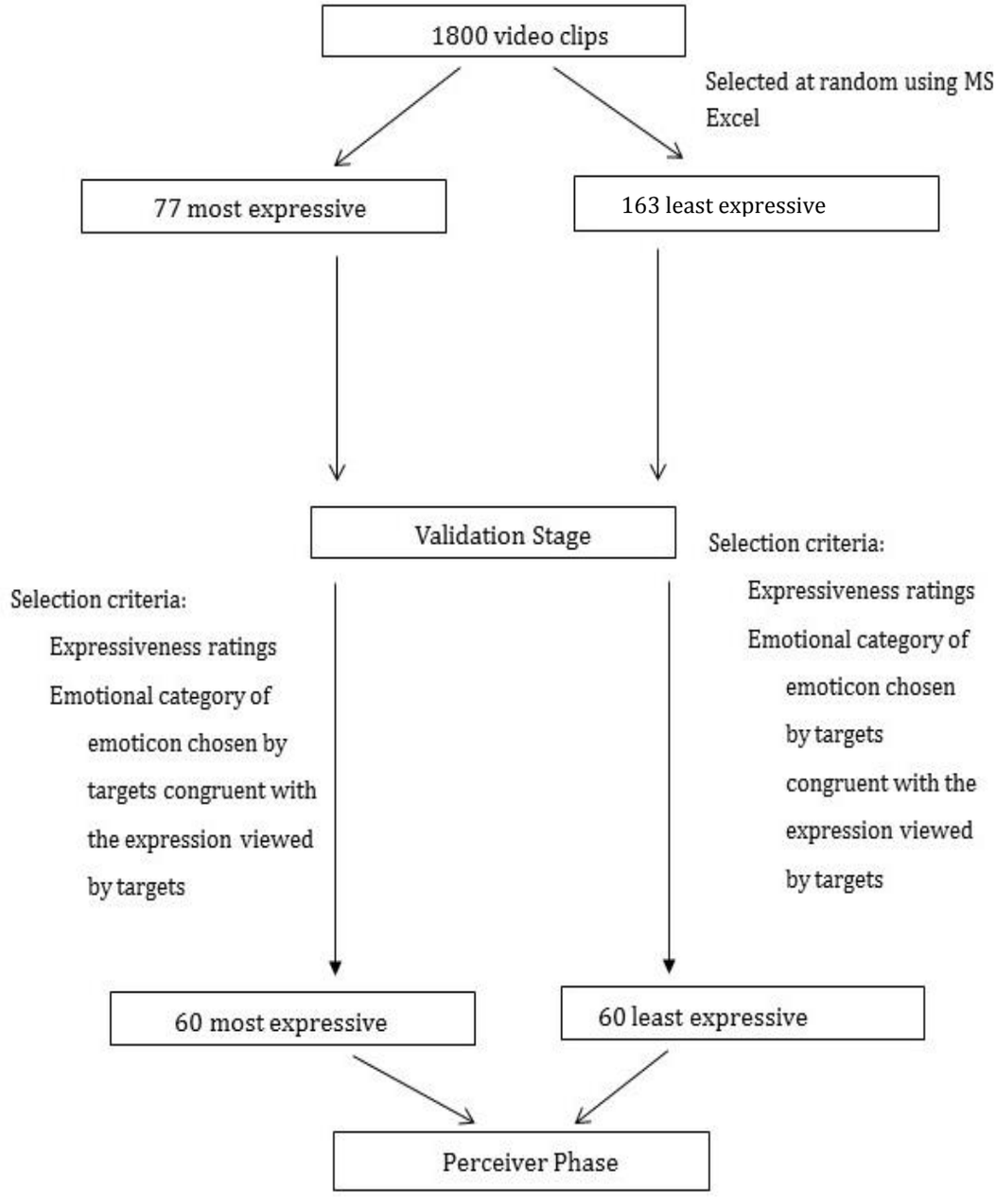


Figure 5.4. The flowchart depicts the selection and validation of stimuli. From a total of 1800 individual video clips, 240 video clips (77 most expressive and 163 least expressive) were selected for the validation stage. From the expressiveness ratings obtained in the validation stage, 120 video clips (60 most expressive and 60 least expressive) video clips were obtained for the perceiver phase.

5.2.3 Perceiver Phase

5.2.3.1 Participants

51 participants (23 females, 28 males, Mean age = 25.05 years, SD = 4.61 years) from the undergraduate and postgraduate student population at the University of Nottingham were recruited as perceivers. All participants were given chocolates as a form of inconvenience allowance and written consent was obtained. Twenty-six perceivers (13 females, 13 males) were allocated to the "classification" group and 25 perceivers (10 females, 15 males) were allocated to the "retrodicting" group.

5.2.3.2 Materials and Apparatus

120 dynamic video clips of targets' reactions to emotional expressions were used as stimuli (refer to 5.5.2 for more information on the selection and validation phase). Of these 120 video clips, 60 displayed female targets and 60 displayed male targets. There were 20 video clips of reactions to each expression viewed by targets (happy, surprised, sad, angry, disgusted, neutral). The size of the video clips was 640 pixels in width and 480 pixels in height. Examples of targets' least expressive and most expressive faces are shown in Figure 5.5 and Figure 5.6 respectively. In addition, there were an equal number of reactions to each expression for both least expressive and most expressive video clips. Although the number of female and male facial expressions was equal for each expression and expressiveness level, some targets appeared repeatedly due to the selection process. The number of unique targets for each expression, separated by expressiveness level, is shown in Table 5.1 and the expressiveness rating of each expression (as rated by judges in the validation stage) is shown in Table 5.5.

[Figure 5.5 was removed to maintain confidentiality]

Figure 5.5. Targets' least expressive reactions to photographs of emotional expressions: (a) happy, (b) surprised, (c) sad, (d) angry, (e) disgusted and (f) neutral.

[Figure 5.6 was removed to maintain confidentiality]

Figure 5.6. Targets' most expressive reactions to photographs of emotional expressions: (a) happy, (b) surprised, (c) sad, (d) angry, (e) disgusted and (f) neutral.

Table 5.1 *The number of unique female and male targets for each expression viewed by targets, separated by expressiveness level.*

Expressiveness	Expression viewed by targets	Number of unique female targets	Number of unique male targets	**Total Number of unique targets**
Least expressive	Happy	5	5	**10**
	Surprised	4	5	**9**
	Sad	3	5	**8**
	Angry	5	5	**10**
	Disgusted	5	5	**10**
	Neutral	3	4	**7**
Most expressive	Happy	5	4	**9**
	Surprised	4	4	**8**
	Sad	3	5	**8**
	Angry	5	3	**8**
	Disgusted	3	4	**7**
	Neutral	5	4	**9**

Table 5.2 *Mean expressiveness ratings of each type of reaction, as rated by judges in the validation stage.*

Expression viewed by targets	Least expressive	Most expressive
Happy	1.60	5.46
Surprised	1.57	5.59
Sad	1.64	5.35
Angry	1.53	5.57
Disgusted	1.66	3.10
Neutral	1.50	5.18

The emoticons used as the response options were identical as the target phase and the order remained the same. The experiment was conducted on PsychoPy v. 1.80 and displayed on a Toshiba Satellite C655-S5501 laptop. The stimuli were randomized using PsychoPy.

5.2.3.3 Design

Perceivers were alternately allocated into two different groups (A between-groups factor of 2 levels- retrodicting, classifying). All participants saw the same expressions (A within-subjects factor of 6 levels- happy, surprised, sad, angry, disgusted, neutral) and the same expressiveness levels (A within-subjects factor of 2 levels- least expressive, most expressive).

5.2.3.4 Procedure

All perceivers were seated approximately 60 cm from the laptop. They were told that they would be looking at videos of various facial expressions. All perceivers began with a practice trial in order to familiarise themselves with the experimental set-up. In the “classification” condition, perceivers were asked to

select the emoticon that best represents the displayed expression. While viewing the target's expression, the question "Which emoticon best represents the displayed expression?" was displayed below. In the "retrodiction" group, perceivers were told that target expressions were reactions to emotional expressions they were viewing and that targets were asked to select the emoticon that best represents the emotional expression they were viewing. Perceivers were then asked to select the emoticon that they thought targets chose. While viewing the target's expression, the question "Which emoticon did this person choose?" was displayed below the expression. All perceivers viewed the same video clips. While targets' expressions were displayed for 3 seconds, each trial was displayed indefinitely until perceivers responded. Perceivers were told to respond as quickly and accurately as possible.

5.3 RESULTS

Data analysis was conducted with SPSS 21.0.

5.3.1 Were perceivers systematically accurate in their responses?

In order to determine whether there was any bias in perceivers' responses, a confusion matrix was created to illustrate the pattern of responses as shown in Table 5.3.

Table 5.3 *Confusion matrix depicting the percentage of responses (and data count) for both groups of perceivers. Shaded cells indicate correct inferences*

		Stimulus Category						
		Happy	Surprised	Sad	Angry	Disgusted	Neutral	Total responses
Classifying Group	Happy	34.81 (181)	5.38 (28)	4.42 (23)	1.92 (10)	13.46 (70)	0.77 (4)	60.77 (316)
	Surprised	3.08 (16)	11.54 (60)	7.12 (37)	6.73 (35)	4.62 (24)	2.31 (12)	35.38 (184)
	Sad	7.69 (40)	13.65 (71)	15.77 (82)	11.15 (58)	9.62 (50)	14.42 (75)	72.31 (376)
	Angry	6.15 (32)	8.27 (43)	13.85 (72)	23.08 (120)	15.19 (79)	12.12 (63)	78.65 (409)
	Disgusted	1.35 (7)	4.62 (24)	5.96 (31)	5.00 (26)	17.69 (92)	3.08 (16)	37.69 (196)
	Neutral	46.92 (244)	56.54 (294)	52.88 (275)	52.12 (271)	39.42 (205)	67.31 (350)	315.19 (1639)
	Total	100.00 (520)	100.00 (520)	100.00 (520)	100.00 (520)	100.00 (520)	100.00 (520)	
Retrodicting Group	Happy	39.40 (197)	9.40 (47)	10.00 (50)	9.20 (46)	14.00 (70)	6.60 (33)	88.60 (443)
	Surprised	8.40 (42)	20.00 (100)	9.40 (47)	11.20 (56)	5.40 (27)	7.60 (38)	62.00 (310)
	Sad	13.20 (66)	17.80 (89)	18.40 (92)	16.40 (82)	10.80 (54)	18.20 (91)	94.80 (474)
	Angry	5.80 (29)	5.20 (26)	12.00 (60)	17.20 (86)	12.80 (64)	10.80 (54)	63.80 (319)
	Disgusted	3.80 (19)	8.40 (42)	8.80 (44)	7.80 (39)	26.60 (133)	4.20 (21)	59.60 (298)
	Neutral	29.40 (147)	39.20 (196)	41.40 (207)	38.20 (191)	30.40 (152)	52.60 (263)	231.20 (1156)
	Total	100.00 (520)	100.00 (520)	100.00 (520)	100.00 (520)	100.00 (520)	100.00 (520)	

Since there was a bias in perceivers' responses (as shown in Table 5.3), unbiased hit rates (H_u) (refer to 1.3.1 for more information) were calculated. The mean H_u was calculated by averaging each perceiver's individual H_u scores. One-

samples *t*-tests were conducted to determine if the mean H_u was significantly above the respective P_c for each type of expressions. Table 5.4 shows the mean H_u, P_c, *t*-value and *p*-values for each category of expression, separated by expressiveness.

Table 5.4 *Classification perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets.*

	Least expressive				**Most expressive**			
Expression viewed by targets	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**
Happy	0.04 (0.06)	0.009	29.60	0.007	0.41 (0.11)	0.034	18.58	<0.001
Surprised	0.01 (0.04)	0.004	0.68	0.580	0.10 (0.08)	0.015	5.37	<0.001
Sad	0.03 (0.03)	0.022	1.04	0.310	0.07 (0.07)	0.019	4.03	<0.001
Angry	0.08 (0.07)	0.014	4.50	0.001	0.08 (0.06)	0.030	4.48	<0.001
Disgusted	0.01 (0.01)	0.004	0.63	0.580	0.20 (0.16)	0.017	6.06	<0.001
Neutral	0.14 (0.05)	0.114	2.42	0.028	0.18 (0.08)	0.061	7.92	<0.001

Table 5.5 *Retrodicting perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets.*

	Least expressive				**Most expressive**			
Expression viewed by targets	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**
Happy	0.06 (0.06)	0.021	4.55	<0.001	0.37 (0.17)	0.028	9.91	<0.001
Surprised	0.03 (0.05)	0.011	2.24	0.040	0.15 (0.10)	0.024	6.25	<0.001
Sad	0.03 (0.03)	0.030	-0.53	0.600	0.07 (0.05)	0.022	4.86	<0.001
Angry	0.05 (0.06)	0.013	3.18	0.004	0.07 (0.06)	0.022	4.02	<0.001
Disgusted	0.03 (0.07)	0.007	1.57	0.130	0.25 (0.16)	0.026	6.85	<0.001
Neutral	0.13 (0.08)	0.085	2.80	0.010	0.14 (0.09)	0.044	5.69	<0.001

5.3.2 Were there any significant group differences in perceivers' judgements between groups?

A 2 x 2 x 6 ANOVA was conducted on H_u with "Group" (retrodicting, classification) as a between-subjects factor, "Expressiveness" (least expressive, most expressive) and "Expression Viewed" (happy, surprised, sad, angry, disgusted, neutral) as within-subjects factors.

Mauchly's test revealed that the assumption of sphericity was violated for the main effect of "Expression Viewed" [$\chi^2(14)=35.98$, $p=0.001$] and interaction between "Expression Viewed" and "Expressiveness" [$\chi^2(14)=42.56$, $p<0.001$]. In these instances, Huynh-Feldt correction was applied.

The results revealed a significant main effect of "Expressiveness", $F(1,49)=213.79$, $p<0.001$, $\eta_p^2=0.81$ indicating that perceivers were significantly more accurate for the most expressive target videos as compared to least expressive target videos. There was also a significant main effect of "Expression Viewed", $F(4.46, 218.59)=68.21$, $p<0.001$, $\eta_p^2=0.58$. Perceivers were significantly more accurate for targets reacting to happy than surprised ($p<0.001$), sad ($p<0.001$), angry ($p<0.001$), disgusted ($p<0.001$) and neutral ($p<0.001$). Reactions to disgusted expressions was significantly more accurate as compared to surprised ($p=0.005$), sad ($p<0.001$), and angry ($p=0.005$) expressions. Perceivers were significantly more accurate for reactions to neutral than surprised ($p<0.001$), sad ($p<0.001$) and angry ($p<0.001$) expressions. But, there was no significant main effect of "Group", $F(1,49)=0.016$, $p=0.90$, $\eta_p^2<0.001$. As for the interaction effects, there was a significant "Expression Viewed" x "Group" interaction, $F(5,245)=2.96$,

p=0.013, η_p^2=0.06 (as shown in Figure 5.7). In addition, there was a significant "Expressiveness" x "Expression Viewed" interaction, F(4.19,205.35)=61.53, p<0.001, η_p^2=0.56 (as shown in Figure 5.8). However, there was no significant "Expressiveness" x "Group" interaction, F(1,49)=0.09, p=0.97, η_p^2=0.002. There was also no significant "Expressiveness" x "Expression Viewed" x "Group" interaction, F(5,245)=0.97, p=0.44, η_p^2=0.02.

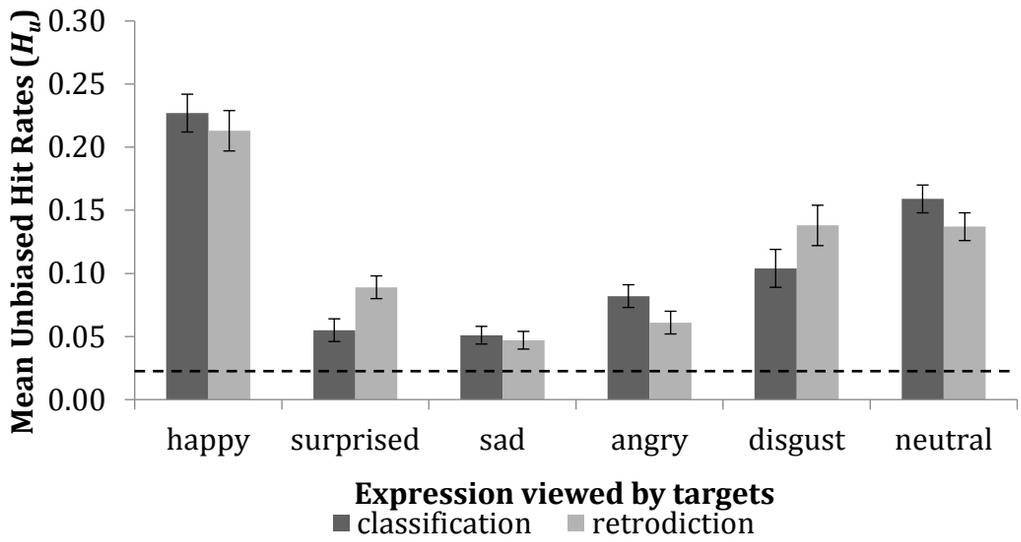


Figure 5.7. Interaction between "Expression Viewed" and "Group". Error bars represent the standard error of mean. The dotted line represents the average chance probability (P_c = 0.03).

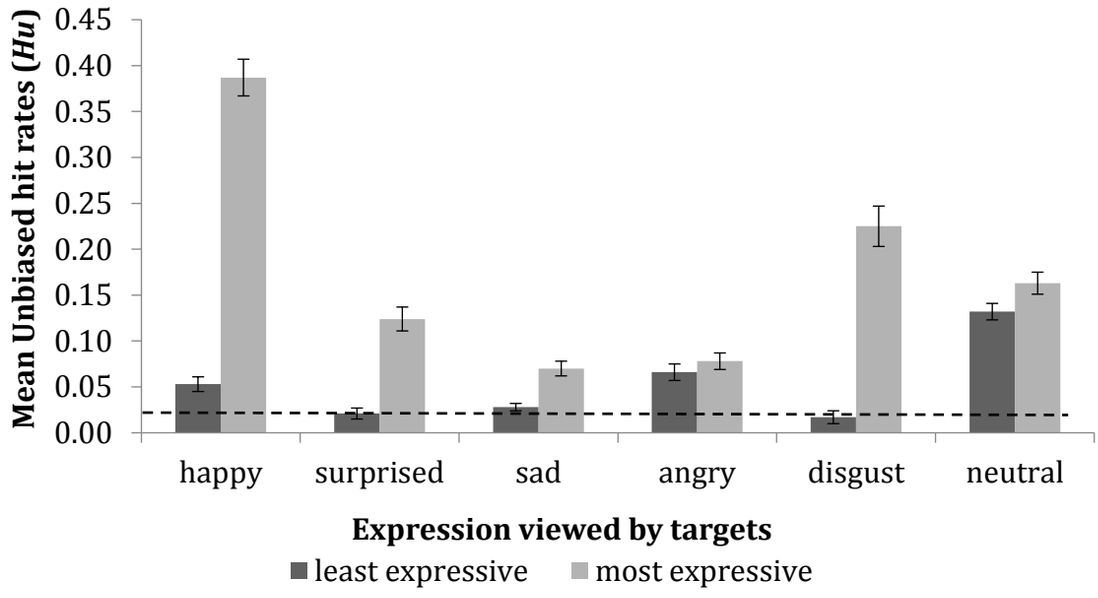


Figure 5.8. Interaction between "Expressiveness" and "Expression Viewed". Error bars represent the standard error of mean. The dotted line represents the average chance probability (P_c =0.03).

5.3.2.1 Post-hoc analyses for "Expression Viewed" x "Group" interaction

Post-hoc analyses (with Bonferroni correction) were conducted to explain the interaction between "Expression Viewed" and "Group". Six independent samples *t*-tests were conducted to identify any group differences in perceiver responses to each expression. There was only a significant difference between groups for surprised, $t(49)=2.83$, $p=0.007$, whereby perceivers in the retrodicting group performed significantly better than perceivers in the classification group.

When perceivers were asked to choose the emoticon which best represents the expression ("classification"), there was a significant main effect of "Expression Viewed", $F(5,125)=47.58$, $p<0.001$, $\eta_p^2=0.66$. Perceivers were significantly more accurate for reactions to happy than surprised ($p<0.001$), sad ($p<0.001$), angry

(p<0.001), disgusted (p<0.001) and neutral (p<0.001) expressions. When targets were reacting to disgusted expressions, perceivers were systematically more accurate than reactions to surprised (p=0.041) and sad (p=0.043) expressions. Perceivers were also significantly more accurate for targets' reactions to neutral than surprised (p<0.001), sad (p<0.001) and angry (p<0.001) expressions. When perceivers were asked to choose the emoticon that they thought targets chose ("retrodicting"), there was a significant main effect of "Expression Viewed", F(4.41, 105.80)=24.50, p<0.001, η_p^2=0.51. Perceivers were significantly more accurate when targets were reacting to happy than surprised (p<0.001), sad (p<0.001), angry (p<0.001), disgusted (p=0.024) and neutral (p=0.014) expressions. Perceivers were also significantly more accurate for reactions to neutral than surprised (p=0.044), sad (p<0.001) and angry (p=0.003) expressions, and performed significantly more accurate for reactions to disgusted than sad (p=0.001) and angry (p=0.006) expressions. In order to illustrate this interaction, Table 5.6 shows the pattern of rank ordering of responses.

Table 5.6 *The pattern of rank order of responses from the highest Mean H_u to the lowest Mean H_u.*

Classifying Perceivers were asked to select an emoticon that represents the target's expression	Retrodicting Perceivers were asked to guess which emoticon the target selected
Happy ↓ (p<0.001) Neutral ↓ (p<0.001) Disgusted ↓ (p<0.001) Angry ↓ (p>0.05) Surprised ↓ (p>0.05) Sad	Happy ↓ (p<0.05) Disgusted ↓ (p>0.05) Neutral ↓ (p<0.05) Surprised ↓ (p>0.05) Angry ↓ (p>0.05) Sad

5.3.2.3 Post-hoc analyses for "Expression Viewed" x "Expressiveness" Interaction

In order to interpret the interaction between "Expressiveness" and "Expression Viewed", post-hoc analyses were conducted. Six independent samples *t*-tests (with Bonferroni correction) were conducted for each level of expression viewed by targets. Perceivers were significantly more accurate for more expressive videos than least expressive videos for happy [$t(50)=16.14$, $p<0.001$], surprised [$t(50)$]6.63, $p<0.001$], sad [$t(50)=4.48$, $p<0.001$], and disgusted [$t(50)=8.53$, $p<0.001$].

For the least expressive videos, there was a significant main effect of "Expression Viewed", $F(4.56, 238.20)=32.51$, $p<0.001$, $\eta_p^2=0.39$. Perceivers were systematically more accurate for reactions to happy than surprised ($p=0.033$) and disgusted ($p=0.021$). Perceivers were also significantly more accurate for reactions to angry than surprised ($p=0.002$), sad ($p=0.013$) and disgusted ($p=0.003$). When targets were reacting to neutral expressions, perceivers were significantly more accurate than reactions to happy ($p<0.001$), surprised ($p<0.001$), sad ($p<0.001$), angry ($p<0.001$) and disgusted ($p<0.001$) expressions. For the most expressive videos, there was a significant main effect of "Expression Viewed", $F(3.80, 190.10)=73.06$, $p<0.001$, $\eta_p^2=0.59$. Perceivers were significantly more accurate for reactions to happy than surprised ($p<0.001$), sad ($p<0.001$), angry ($p<0.001$), disgusted ($p<0.001$) and neutral ($p<0.001$). Perceivers were significantly more accurate for reactions to surprised expressions than sad ($p=0.021$) while perceivers were significantly more accurate for reactions to disgusted expressions than surprised ($p=0.003$), sad ($p<0.001$) and angry ($p<0.001$) expressions. Lastly, perceivers were also significantly more accurate when they viewed targets reacting to neutral than sad ($p<0.001$) and angry ($p<0.001$) expressions.

5.3.3 Were there any significant group differences in perceivers' response times?

Firstly, outliers were identified using the Median Absolute Deviation (MAD) (Leys et al., 2013) method where values which fall below and above 2.5 standard deviation are excluded from the analyses. Fifteen out of 612 individual cases (2.3%) were excluded from the analyses. The tests of normality showed that the

assumption of normality was not violated for all variables in both perceiver groups, *Shapiro-Wilk*, $p>0.05$.

The means and standard deviations for both groups (excluding outliers), across different expressions and different levels of expressiveness are shown in Table 5.7.

Table 5.7 *Mean (and standard deviation) response times (in ms) for both perceiver groups.*

	Classifying perceivers		**Retrodicting perceivers**	
Expression Viewed	**Least expressive**	**Most expressive**	**Least expressive**	**Most expressive**
Happy	4397.28 (753.05)	4552.09 (779.29)	5617.30 (1181.50)	5018.75 (1277.63)
Surprised	4545.48 (794.31)	5215.82 (986.81)	5407.87 (1828.68)	5422.82 (1163.90)
Sad	4663.60 (986.48)	4833.67 (809.59)	5667.52 (1226.67)	5613.48 (1791.41)
Angry	4667.86 (790.46)	4627.37 (1318.89)	5742.06 (978.75)	5305.21 (1133.47)
Disgusted	4718.80 (1047.17)	4962.83 (958.29)	5697.04 (1567.53)	5444.98 (1316.16)
Neutral	4410.86 (910.45)	4336.48 (684.45)	5633.86 (1012.24)	5246.84 (1447.39)

A 2 x 6 x 2 mixed ANOVA was conducted with "Expressiveness" (least expressive, most expressive), "Expression Viewed" (happy, surprised, sad, angry, disgusted, neutral) as within-subjects factors and "Group" (retrodicting, classifying) as between-subjects factors. Mauchly's test of sphericity revealed that the assumption

of sphericity was violated for the main effect of "Expression Viewed" [$\chi^2(14)=34.28$, $p=0.002$], thus, Huynh-Feldt correction was applied.

There was a significant main effect of "Expression Viewed", $F(4.30, 189.31)=3.78$, $p=0.005$, $\eta_p^2=0.008$, where perceivers took significantly more time to respond to disgusted reactions than happy ($p=0.004$) and neutral ($p=0.003$) reactions. There was also a significant main effect of "Group", $F(1,44)=9.99$, $p=0.003$, $\eta_p^2=0.19$ whereby retrodicting perceivers took significantly more time to respond than classifying perceivers. There was no significant main effect of "Expressiveness", $F(1,44)=0.095$, $p=0.76$, $\eta_p^2=0.002$. As for interactions, there was a significant interaction between "Expressiveness" and "Group", $F(1,44)=8.70$, $p=0.005$, $\eta_p^2=0.17$ (Figure 5.9). There was also a significant interaction between "Expressiveness" and "Expression Viewed", $F(4.70, 206.97)=3.19$, $p=0.01$, $\eta_p^2=0.07$ (Figure 5.10). There was, however, no significant interaction between "Expression Viewed" and "Group", $F(4.30, 189.31)=1.88$, $p=0.11$, $\eta_p^2=0.41$ and no significant interaction between "Expressiveness", "Expression Viewed", and "Group", $F(4.70, 206.97)=0.29$, $p=0.91$, $\eta_p^2=0.006$.

5.3.3.1. Post-hoc analyses for "Expressiveness" x "Group" interaction

In order to interpret this interaction, two independent samples *t*-tests with Bonferroni correction was conducted for each level of "Expressiveness" and two paired samples *t*-tests were conducted for each perceiver group. Retrodicting perceivers took a longer time to respond than classifying perceivers

for least expressive reactions, $t(49)=3.55$, $p=0.001$, and most expressive expressions, $t(49)=2.34$, $p=0.02$. Classifying perceivers took a longer time to respond to the most expressive reactions as compared to the least expressive reactions, $t(25)=3.35$, $p=0.003$ but there was no significant difference in the time taken to respond to most expressive and least expressive reactions for retrodicting perceivers, $t(24)=0.73$, $p=0.48$.

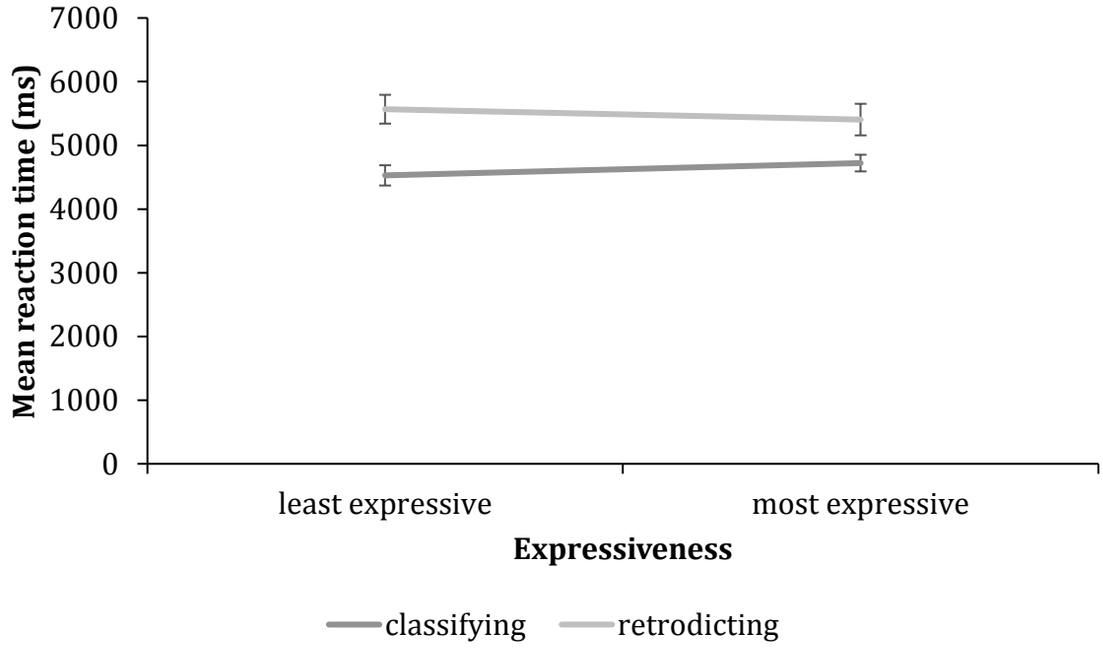


Figure 5.9. Interaction between "Expressiveness" and "Group". Error bars represent the standard error of mean.

5.3.3.2. Post-hoc for "Expressiveness" x "Expression Viewed" interaction

Six paired-samples *t*-tests with Bonferroni correction were also carried out. Perceivers took significantly more time to respond to most expressive reactions than least expressive reactions only for surprised, $t(48)=3.61$, $p=0.001$ but not all

other expressions. Two separate one-way ANOVAs were conducted to investigate the effect of "Expression Viewed" on each level of "Expressiveness". For the least expressive expressions, there was no significant main effect of "Expression Viewed", $F(4.20, 201.72)=1.61$, $p=0.17$, $\eta_p^2=0.03$. For the most expressive reactions, there was a significant main effect of "Expression Viewed", $F(4.41, 202.97)=4.48$, $p=0.001$, $\eta_p^2=0.09$. [Mauchly's test of sphericity, $\chi^2(14)=25.71$, $p=0.028$]. Perceivers took significantly less time to respond to happy than surprised reactions ($p=0.008$). Perceivers also took significantly less time to respond to neutral reactions than surprised ($p<0.001$) and disgusted reactions ($p=0.035$).

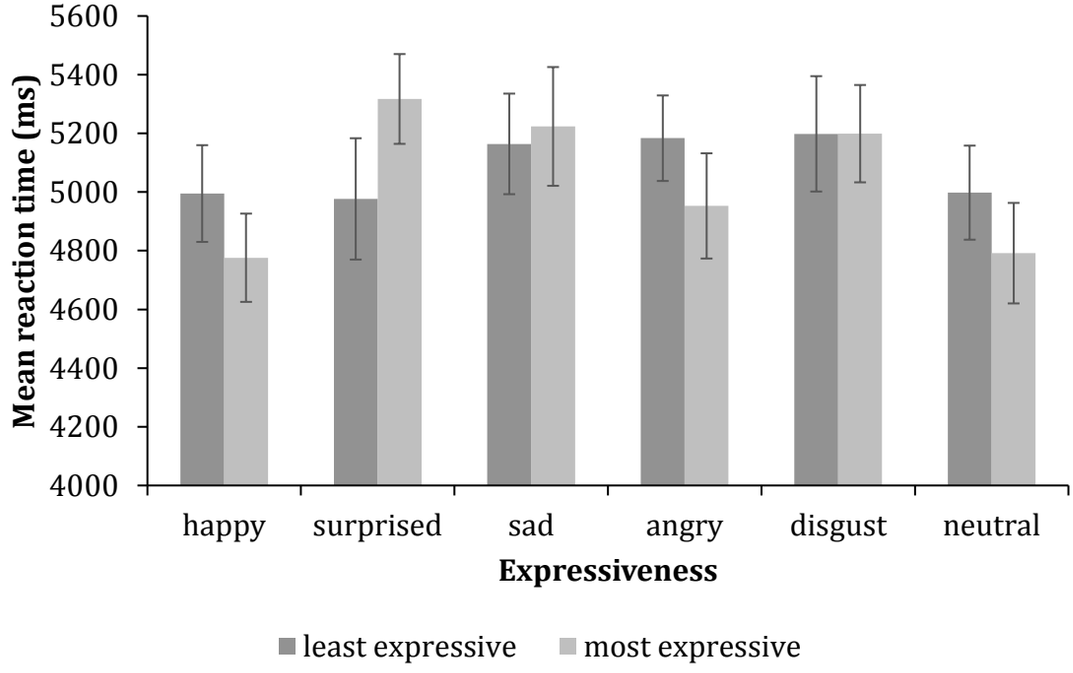


Figure 5.10. *Interaction between "Expressiveness" and "Expression viewed". Error bars represent the standard error of mean.*

5.4 DISCUSSION

The results showed that perceivers in both groups were significantly above chance in their responses, presumably because the emotional expressions that targets were viewing were signalled in targets' expressions. Consequently, these emotional states provided clues which aided perceivers' judgements of which emoticons targets chose. Thus, this accurate inference of which emoticons targets chose also indicates that perceivers were systematically accurate, albeit slightly better than chance, at inferring the emotional expressions viewed by targets because the emotional category of emoticons selected by targets always corresponded with the expressions viewed. Although one may suggest that retrodiction perceivers were merely classifying their responses through a parallel route of processing, they performed differently than classifying perceivers, as seen from the pattern of rank responses. Even if a different pattern of rank responses may not necessarily indicate that a more elaborated strategy is involved, retrodiction perceivers took a longer time to respond than classifying perceivers suggesting that more elaborated processing was used.

Thus, the findings from this experiment suggest that people are above chance level in their retrodictive judgements, suggesting that they are capable of inferring the emotional expressions that targets were viewing. When making such judgements, retrodicting perhaps perceivers were genuinely thinking about targets inner states through an elaborated and indirect route of processing, rather than a simple and direct route used by classifying perceivers.

EXPERIMENT 2:2

5.5 INTRODUCTION

Facial expressions provide a vast range of information that we can draw upon to guide our social interactions, such as basic and complex emotions (Baron-Cohen et al., 1996, 1997; Ekman et al., 1969, 1972) and more recently, situational antecedents of inner states (Cassidy et al., 2014; Pillai et al., 2012; Teoh et al., 2017). When looking at a face for social cues, individuals are attuned to the features of the face such as the eyes and mouth (Henderson et al., 2005; Yarbus, 1967) which are the primary and secondary source of information respectively (Davies et al., 1977; Fraser et al., 1990; Tanaka and Farah, 1993). Drawing upon the suggestion that the eyes are the primary source of information, it has been proposed that the eyes are crucial for understanding inner states (Baron-Cohen et al., 1997). In an earlier study conducted by Argyle (1978), people tended to focus on the eyes of another while looking at the direction of another person's face. Not only that, individuals also spent significantly longer looking at the eyes when looking at faces of emotional expressions from static images (Hernandez et al., 2009; Janik et al., 1978) and dynamic videos (Klin et al., 2002).

However, more recent studies have shown that the mouth region is central in communicating information about inner states (Cunningham et al., 2004), as indicated by an increase in fixation time in the mouth region as compared to the eyes (Kirchner et al., 2011). Others have demonstrated that attention to facial features varies across different types of emotions. For instance, equal attention is given to both eyes and mouth for fearful and neutral expressions whereas attention

to the eyes is significantly more than the mouth for sad and angry expressions (Eisenbarth and Alpers, 2011).

When perceivers were asked to retrodict from targets' behavior to a situational antecedent, perceivers' eye movements varied across different scenarios (Pillai et al., 2012). Individuals looked significantly longer at the eyes for target reactions when they were kept waiting, compared to when they were told a joke or a story. On the other hand, individuals looked significantly longer at the mouth for targets' reactions to joke, story and compliment, as compared to when targets were kept waiting. Although it is not clear as to which facial features are important in understanding inner states, these findings consistently demonstrated attention to the eyes and mouth during these tasks.

This study, therefore, aims to investigate the pattern of eye movements of retrodicting perceivers who will have to infer which emoticons targets chose (similar to Experiment 2:1). Additionally, this study also aims to investigate whether the pattern of eye movements of retrodicting perceivers is significantly different from that of classifying perceivers. Similar to previous studies (Eisenbarth and Alpers, 2011; Kirchner et al., 2011; Pillai et al., 2012), average gaze duration at the eyes and mouth were measured. An increase in gaze duration could reflect more complex decision making, increased focused attention, greater memory loads (Goldberg et al., 2000) and relating the task to internalized representations (Goldberg and Kotval, 1999). As another measure of disentangling between the various processes used in these two groups, a time-course analysis of perceivers' pattern of eye movements over the entire duration of stimulus presentation (3 seconds) split into1.0 second intervals (similar to Freeth et al.

(2010)'s study) was conducted to investigate if there are any significant differences in the pattern of gaze duration at the eyes and mouth between the two perceiver groups.

5.6 METHODS

Ethical approval was obtained from the Ethics Committee, School of Psychology, University of Nottingham.

5.6.1 Participants

46 participants (28 females, 18 females, *Mean age* = 21.98 years, *SD* = 2.70 years) from the undergraduate and postgraduate student population at the University of Nottingham were recruited as perceivers. All participants were given £3 as a form of inconvenience allowance and written consent was obtained. 23 perceivers (13 females, 10 males) were allocated randomly to the "retrodicting" group while 23 perceivers (15 females, 8 males) were allocated randomly to the "classification" group. All participants had normal or corrected vision.

5.6.2 Materials and Apparatus

The stimuli used in this study were obtained from the target's video clips in the previous study (See 5.2.1). In order to avoid any prolonged discomfort due to the eye-tracking head gear, only 72 videos were selected. Based on the ratings of expressiveness obtained from the validation stage (See 5.2.2.1), the 6 least and 6 most expressive video clips for each type of reaction were chosen. Of these 72 video clips, 36 had female targets and another 36 video clips with male targets. The number of unique targets for each type of reaction according to each

expressiveness level is shown in Table 5.8 while their respective expressiveness ratings appear in Table 5.9. The experiment was displayed on SR Research Experiment Builder v. 1.3.40. The size of the video clips was 640 pixels in width and 480 pixels in height. The size of the emoticons was 192 pixels in width and 122 pixels in height. Eye movements were recorded using EyeLink 2.0. The eye-tracking head gear was connected to the display computer and EyeLink Host PC. Eye movements were tracked at 500 samples/second. Where pupil-corneal reflection tracking mode was not possible, only pupil tracking mode was used. Since pupil tracking mode is more susceptible to errors caused by the headband slipping, drift correction was applied before each trial to adjust for any discrepancies. Calibration was performed using a 9-point calibration and these calibration points appeared at random.

Table 5.8 *Number of unique targets for each type of reaction, according to each expressiveness level.*

	Expression Viewed	**Number of unique female targets**	**Number of unique male targets**	**Total unique targets**
least	Happy	3	3	6
	Surprised	3	3	6
	Sad	3	3	6
	Angry	3	3	6
	Disgusted	3	3	6
	Neutral	3	3	6
most	Happy	3	3	6
	Surprised	3	3	6
	Sad	2	3	5
	Angry	3	1	4
	Disgusted	1	3	4
	Neutral	3	2	5

Table 5.9 *Mean expressiveness ratings for each type of reaction.*

Expressiveness	**Expression Viewed**	**Expressiveness ratings**
Least	Happy	1.57
	Surprised	1.52
	Sad	1.58
	Angry	1.48
	Disgusted	1.62
	Neutral	1.40
Most	Happy	2.68
	Surprised	2.77
	Sad	2.85
	Angry	2.77
	Disgusted	3.52
	Neutral	2.28

5.6.3 Procedure

All perceivers sat approximately 60 cm from the display computer screen and were told that they would be looking at several facial expressions. In the classification group, perceivers were told that they had to select the emoticon which best represents the target expression. Perceivers in the retrodicting group were told that these targets were looking at a photograph of an expression and selected an emoticon that represented the expression. Their task was to guess the emoticon that targets chose. Before perceivers started the experiment, their eye movements were calibrated. Then, they began with a practice trial in which perceivers saw target reactions displayed at the centre of the screen while the question and emoticons appeared below the expression at the same time. Targets' reactions were only displayed for three seconds while each trial lasted for ten

seconds. While the perceivers were viewing the expression, the question "Which emoticon best represents the displayed expression?" appeared in the classification condition and "Which emoticon did this person choose?" appeared in the retrodicting condition. All perceivers viewed the same expressions. The numbers which corresponded with the options were displayed below the emoticon. They were asked to respond by pressing "1" for happy, "2" for surprised, "3" for sad, "4" for angry, "5" for disgusted and "6" for the neutral emoticon. Perceivers were told to respond as accurately and as quickly as possible. The experimental set-up is shown in Figure 5.11 and Figure 5.12. After the practice trial, perceivers started with the test trials.

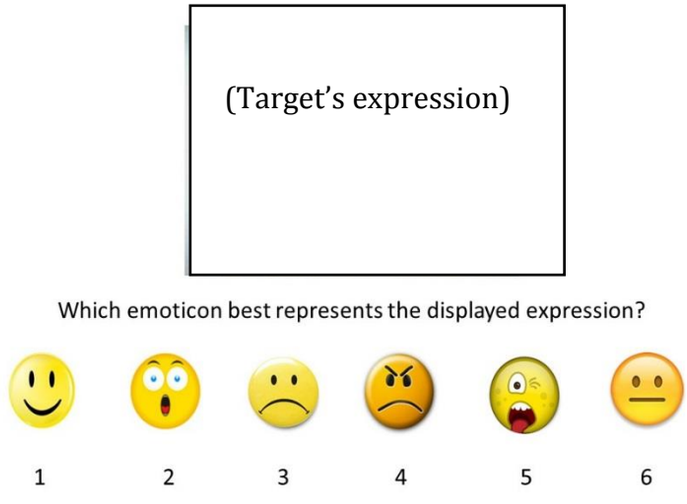


Figure 5.11. Example of an experimental set-up, as seen by classification perceivers.

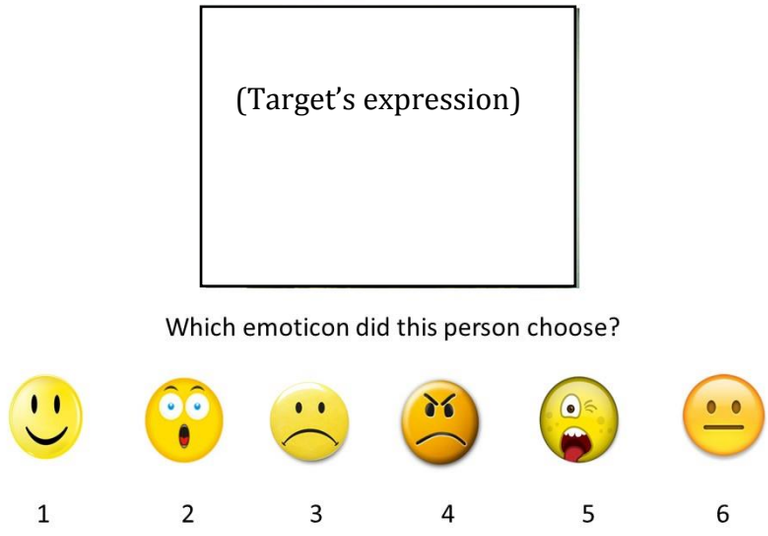


Figure 5.12. Example of an experimental set-up, as seen by retrodicting perceivers.

5.7 RESULTS

The data were analysed using SPSS 22.0. Since each trial terminated after 10 seconds, some perceivers did not respond on some trials, resulting in missing data.

5.7.1 Behavioural data

5.7.1.1 Were perceivers systematically accurate in their responses?

In order to determine if there was any bias in perceivers' responses, a confusion matrix was created, as shown in Table 5.10.

Table 5.10 Confusion matrix depicting the percentage of responses (and data count) for each group of perceivers. Shaded cells indicate correct inferences. The total does not amount to 100.00 due to missing data.

		Stimulus Category						
		Happy	Surprised	Sad	Angry	Disgusted	Neutral	Total responses
"Classification" Group	Happy	36.96 (102)	11.23 (31)	5.43 (15)	3.62 (10)	4.35 (12)	2.90 (8)	64.49 (178)
	Surprised	2.90 (8)	19.57 (54)	15.58 (43)	7.61 (21)	9.06 (25)	5.80 (16)	60.52 (167)
	Sad	12.32 (34)	11.59 (32)	18.84 (52)	17.03 (47)	9.78 (27)	18.84 (52)	88.40 (244)
	Angry	7.97 (22)	6.88 (19)	19.20 (53)	20.65 (57)	15.22 (42)	11.59 (32)	81.51 (225)
	Disgusted	2.90 (8)	2.57 (7)	8.70 (24)	3.26 (9)	28.62 (79)	4.71 (13)	50.73 (140)
	Neutral	35.14 (97)	45.29 (125)	30.43 (84)	46.21 (127)	32.25 (89)	43.62 (148)	242.74 (670)
	Total	98.19 (271)	97.10 (268)	98.18 (271)	98.18 (271)	99.28 (274)	97.46 (269)	
"Retrodicting" Group	Happy	43.48 (120)	19.57 (54)	10.86 (30)	7.61 (21)	13.77 (38)	6.88 (19)	102.17 (282)
	Surprised	5.07 (14)	20.29 (56)	14.13 (39)	5.80 (16)	5.07 (14)	8.33 (23)	58.69 (162)
	Sad	11.59 (32)	13.04 (36)	18.12 (50)	17.39 (48)	13.77 (38)	23.91 (66)	97.82 (270)
	Angry	5.43 (15)	4.71 (13)	15.94 (44)	17.03 (47)	15.94 (44)	9.06 (25)	68.11 (188)
	Disgusted	2.17 (6)	4.35 (12)	14.49 (40)	7.25 (20)	26.09 (72)	5.43 (15)	59.78 (165)
	Neutral	30.07 (83)	35.13 (97)	23.91 (66)	42.39 (117)	22.82 (63)	44.93 (124)	199.25 (550)
	Total	97.81 (270)	97.10 (268)	97.46 (269)	97.47 (269)	97.46 (269)	98.55 (272)	

As there was a bias towards neutral responses, unbiased hit rates (H_u) (refer to 1.3.1 for more information) were calculated to determine if perceivers were systematically accurate in their responses. One sample *t*-tests were carried out to compare Mean H_u against their respective P_c. Table 5.11 and 5.12 show the Mean H_u, P_c, *t*-value and *p*-value for each stimulus for each level of expressiveness separated by group.

Table 5.11 *Classification perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets.*

	Least Expressive				**Most Expressive**			
Expression viewed by targets	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**
Happy	0.056 (0.09)	0.011	2.47	0.02	0.523 (0.20)	0.027	11.71	<0.001
Surprised	0.023 (0.05)	0.011	1.24	0.24	0.178 (0.12)	0.022	6.36	<0.001
Sad	0.046 (0.05)	0.025	2.03	0.06	0.082 (0.09)	0.023	3.30	0.003
Angry	0.059 (0.08)	0.013	2.75	0.01	0.109 (0.102)	0.031	3.66	0.001
Disgusted	0.004 (0.02)	0.002	0.45	0.55	0.378 (0.205)	0.026	8.25	<0.001
Neutral	0.123 (0.06)	0.096	2.43	0.03	0.170 (0.103)	0.034	6.37	<0.001

Table 5.12 *Retrodicting perceivers' mean H_u (and standard deviation) and P_c for each category of expressions viewed by targets.*

	Least Expressive				**Most Expressive**			
Expression viewed by targets	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**	**Mean H_u (and standard deviation)**	**P_c**	***t*-value**	***p*-value**
Happy	0.040 (0.07)	0.024	1.14	0.27	0.498 (0.175)	0.031	12.83	<0.001
Surprised	0.017 (0.04)	0.009	1.05	0.31	0.169 (0.114)	0.023	6.13	<0.001
Sad	0.064 (0.08)	0.028	2.22	0.04	0.045 (0.04)	0.025	2.32	0.03
Angry	0.036 (0.08)	0.010	1.62	0.12	0.098 (0.09)	0.027	5.48	<0.001
Disgusted	0.021 (0.06)	0.004	1.36	0.19	0.278 (0.22)	0.028	5.48	<0.001
Neutral	0.110 (0.08)	0.085	1.57	0.13	0.149 (0.12)	0.024	5.12	<0.001

5.7.1.2. Were there any significant differences in perceivers' judgements between groups?

A 6 x 2 x 2 ANOVA was conducted on Mean H_u with "Expression Viewed" (happy, surprised, sad, angry, disgusted, neutral), "Expressiveness" (least expressive, most expressive) as within-subjects factors and "Group" (retrodicting, classifying) as a between-subjects factor.

Mauchly's test of sphericity revealed that the assumption of sphericity was violated for the main effect of "Expression Viewed" [$\chi^2(14)=40.35$, $p<0.001$] and interaction between "Expressiveness" and "Expression Viewed" [$\chi^2(14)=39.74$, $p<0.001$]. When the assumption of sphericity was violated, Huynh-Feldt correction was applied.

There was a significant main effect of "Expressiveness", $F(1,44)=280.43$, $p<0.001$, $\eta_p^2=0.86$, suggesting that perceivers were more accurate for most expressive than least expressive stimuli. There was also a significant main effect of "Expression Viewed", $F(4.04, 177.77)=53.18$, $p<0.001$, in which pairwise comparisons (with Bonferroni correction) showed that reactions to happy was recognized more accurately as compared to surprised ($p<0.001$), sad ($p<0.001$), angry ($p<0.001$), disgusted ($p<0.001$), and neutral ($p<0.001$). Reactions to surprised expressions were also recognized more accurately as compared to sad ($p=0.036$). Perceivers were more accurate in recognizing targets' reactions to disgusted than surprised ($p=0.005$), sad ($p<0.001$) and angry ($p<0.001$) expressions. Lastly, reactions to neutral expressions were more accurate than reactions to surprised ($p=0.017$), sad ($p<0.001$) and angry ($p<0.001$). But, there was no significant main effect of group, $F(1,44)=3.51$, $p=0.068$, $\eta_p^2=0.074$. There

was a significant interaction between "Expressiveness" and "Expression Viewed", $F(4.17, 183.35)=63.26$, $p<0.001$, $\eta_p^2=0.59$. There was no interaction between "Expressiveness" and "Group", $F(1,44)=2.02$, $p=0.16$, $\eta_p^2=0.044$, no interaction between "Expression Viewed" and "Group", $F(4.04, 177.77)=0.30$, $p=0.88$, and no significant interaction between "Expressiveness", "Expression Viewed" and "Group", $F(4.17, 183.35)=1.14$, $p=0.34$, $\eta_p^2=0.025$.

5.7.1.2.1 Post-hoc analyses for "Expressiveness" and "Expression Viewed" interaction

Six paired samples *t*-tests (with Bonferroni correction) showed that perceivers were significantly more accurate for most expressive stimuli as compared to least expressive stimuli for the following types of reactions: happy, $t(45)=16.05$, $p<0.001$; surprised, $t(45)=8.36$, $p<0.001$; angry, $t(45)=2.86$, $p=0.006$ and disgusted, $t(45)=9.63$, $p<0.001$. Two separate one-way ANOVAs were conducted for each level of expressiveness. For the least expressive stimuli, there was a significant main effect of "Expression Viewed", $F(4.69, 210.83)=16.13$, $p<0.001$, $\eta_p^2=0.26$ [Assumption of sphericity was violated, $\chi^2(14)=24.35$, $p=0.042$]. Post-hoc with Bonferroni correction revealed that reactions to sad expressions was recognized more accurately than disgusted ($p=0.005$) while reactions to neutral expressions was recognized more accurately than happy ($p<0.001$), surprised ($p<0.001$), sad ($p<0.001$), angry ($p=0.001$) and disgusted ($p<0.001$) expressions. As for the most expressive stimuli, there was a significant main effect of "Expression Viewed", $F(3.58, 160.95)=67.92$, $p<0.001$, $\eta_p^2=0.60$ [The assumption of sphericity was violated, $\chi^2(14)=59.56$, $p<0.001$]. When targets were reacting to

happy expressions, perceivers were significantly more accurate than when targets reacted to surprised (p<0.001), sad (p<0.001), angry (p<0.001), disgusted (p=0.001) and neutral (p<0.001). Reactions to surprised expressions were recognized significantly more accurately than sad (p<0.001) and angry (p=0.046). Reactions to disgusted expressions was also recognized significantly more than surprised (p=0.002), sad (p<0.001), angry (p<0.001) and neutral (p<0.001). Lastly, perceivers were systematically more accurate in recognizing reactions to neutral than sad (p<0.001). The graph depicting the interaction between "Expressiveness" and "Expression Viewed" can be seen in Figure 5.13.

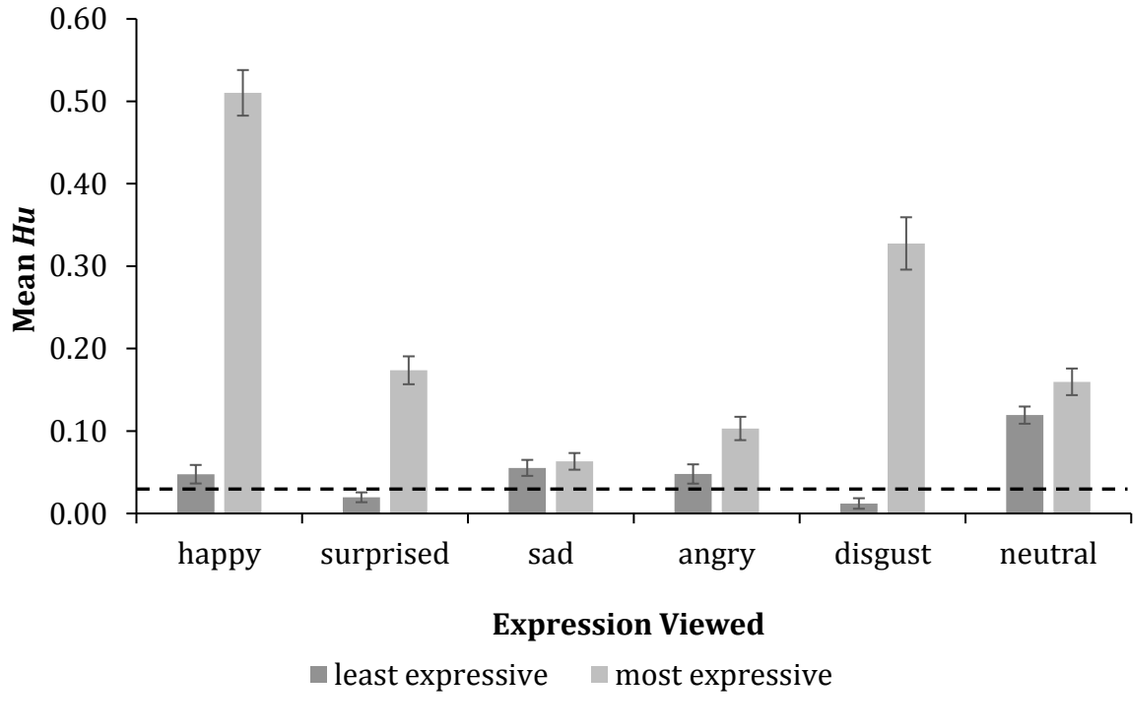


Figure 5.13. The interaction between the expressions viewed by targets and their expressiveness level. The dotted line represents the average chance probability (*Pc* = 0.03). Error bars represent standard error of mean.

5.7.1.3 Were there any significant group differences in perceivers' response times?

Outliers were identified using Median Absolute Deviation (MAD) (Leys et al., 2013) and 7 out of 552 individual cases (1.23%) was removed. The assumption of normality was not violated for both groups, *Shapiro-Wilk, p*>0.05.

Mean and standard deviation for response times, without outliers, in both groups can be seen from Table 5.13.

Table 5.13 *Mean (and standard deviation) response times (in ms) for both perceiver groups.*

	Classifying group		**Retrodicting group**	
Expression Viewed	**Least expressive**	**Most expressive**	**Least expressive**	**Most expressive**
Happy	4524.15 (738.55)	4189.18 (955.87)	4851.99 (768.62)	3872.00 (802.91)
Surprised	4599.15 (698.00)	4614.96 (1054.91)	4584.05 (726.08)	4608.95 (1191.11)
Sad	4406.41 (687.74)	4251.97 (792.49)	4786.76 (846.28)	4495.73 (879.95)
Angry	4589.42 (886.16)	4628.75 (962.07)	5003.49 (875.89)	4822.11 (732.25)
Disgusted	4474.15 (773.62)	4756.54 (684.12)	4640.54 (774.10)	4943.87 (718.01)
Neutral	4546.47 (814.05)	4634.87 (717.60)	4837.93 (900.69)	4833.18 (1022.99)

A 6 x 2 x 2 ANOVA was conducted on response time with "Expression Viewed" (happy, surprised, sad, angry, disgusted, neutral), "Expressiveness" (least

expressive, most expressive) as within-subjects factors and "Group" (retrodicting, classifying) as a between-subjects factor. There was a significant main effect of "Expressiveness", $F(1,41)=4.72$, $p=0.04$, $\eta_p^2=0.10$ whereby perceivers took significantly more time to respond to least expressive stimuli than most expressive stimuli. There was also a significant main effect of "Expression Viewed", $F(5,205)=9.79$, $p<0.001$, $\eta_p^2=0.19$. Perceivers took significantly less time to respond to happy than surprised ($p=0.002$), angry ($p<0.001$), disgusted ($p<0.001$) and neutral ($p<0.001$). Perceivers also took significantly less time to respond to sad than neutral ($p=0.022$). There was, however, no significant main effect of "Group", $F(1, 41)=0.37$, $p=0.55$, $\eta_p^2=0.009$. As for interactions, there was a significant interaction between "Expressions Viewed" and "Group", $F(5,205)=2.27$, $p=0.05$, $\eta_p^2=0.05$ (as shown in Figure 5.14). There was also a significant interaction between "Expressiveness" and "Expressions Viewed", $F(5,205)=8.07$, $p<0.001$, $\eta_p^2=0.17$ (as shown in Figure 5.15). However, there was no significant interaction between "Expressiveness" and "Group", $F(1,41)=2.77$, $p=0.10$, $\eta_p^2=0.06$, and no significant interaction between "Expressiveness", "Expressions Viewed" and "Group", $F(5, 205)=1.19$, $p=0.31$, $\eta_p^2=0.028$.

5.7.1.3.1 Post-hoc for interaction between "Expressions Viewed" and "Group"

Six paired samples *t*-tests (with Bonferroni correction) showed that there were no significant group differences in time taken to respond to all types of reactions, $p>0.05$. Two separate one-way ANOVAs were conducted for each

perceiver group. For retrodicting perceivers, there was a significant main effect of "Expressions Viewed", $F(5,110)=10.60$, $p<0.001$, $\eta_p^2=0.33$. Perceivers took significantly less time to respond to happy than all other reactions [surprised ($p=0.003$); angry ($p<0.001$); disgusted ($p<0.001$) and neutral ($p=0.001$). As for classifying perceivers, there was a significant main effect of "Expressions Viewed", $F(5,110)=3.03$, $p=0.01$, $\eta_p^2=0.12$. But, post-hoc tests did not reveal any significant differences in reaction times between different reactions.

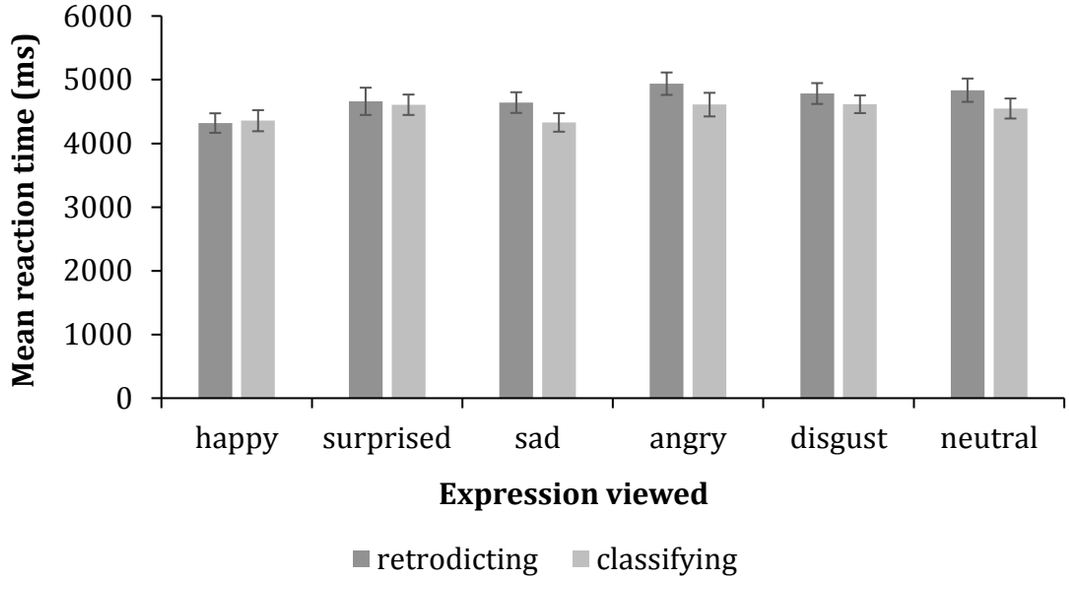


Figure 5.14. Interaction between "Expressions Viewed" and "Group". Error bars represent the standard error of mean.

Table 5.14 Rank ordering of Reaction times (RT) for the various expressions in both groups, from the fastest (top) to the slowest (bottom)

Classifying Perceivers were asked to select an emoticon that represents the target's expression	Retrodicting Perceivers were asked to guess which emoticon the target selected
Sad ↓ (p>0.05) Happy ↓ (p>0.05) Neutral ↓ (p>0.05) Surprised ↓ (p>0.05) Angry ↓ (p>0.05) Disgusted	Happy ↓ (p=0.04) Sad ↓ (p>0.05) Surprised ↓ (p>0.05) Disgusted ↓ (p>0.05) Neutral ↓ (p>0.05) Angry

5.7.1.3.2 Post-hoc for interaction between "Expressiveness" and "Expressions Viewed"

Six paired-samples *t*-tests with Bonferroni Correction showed that perceivers responded significantly faster for most expressive expressions than the least expressive expressions, only for happy reactions, $t(44)=-5.60$, $p<0.001$. Two separate one-way ANOVAs were conducted for each level of "Expressiveness". For least expressive expressions, there was no significant main effect of "Expressions

Viewed", $F(5, 215)=0.80$, $p=0.55$, $\eta_p^2=0.02$. There was, however, a significant main effect of "Expressions Viewed" for the most expressive stimuli, $F(5, 210)=12.92$, $p<0.001$, $\eta_p^2=0.24$ where perceivers took significantly less time to respond to happy than surprised, sad ($p=0.011$), angry ($p<0.001$), disgusted ($p<0.001$) and neutral ($p<0.001$). Perceivers also took significantly less time to respond to sad than angry ($p=0.001$), disgusted ($p<0.001$) and neutral ($p=0.001$).

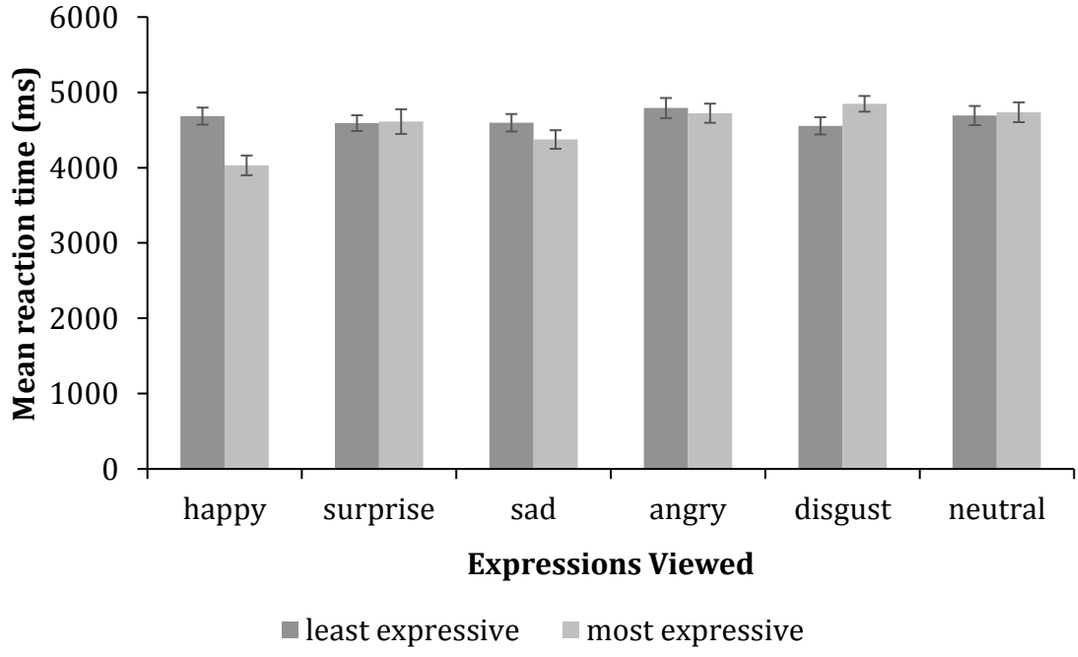


Figure 5.15. Interaction between "Expressions Viewed" and "Expressiveness". Error bars represent the standard error of mean.

5.7.2 Eye-tracking data

5.7.2.1 Group differences in the overall pattern of eye movements

Two separate 6 x 2 x 2 ANOVAs were conducted for each area of interest (eyes, mouth) on the average gaze duration (ms), with "Expression Viewed"

(happy, surprised, sad, angry, disgusted, neutral) and "Expressiveness" (least expressive, most expressive) as within-subjects factors while "Group" (classifying, retrodicting) was a between-subjects factor.

5.7.2.1.1. Eyes

There were no outliers as indicated by the Median Absolute Deviation (Leys et al., 2013) and the assumption of normality was not violated for all variables for both groups, *Shapiro-Wilks, p*>0.05. The means and standard deviation for each type of reaction is shown in Table 5.15.

Table 5.15. *Mean (and standard deviation) of gaze duration at the eyes for each type of reaction, separated by each perceiver group.*

	Classifying		**Retrodicting**	
	Least expressive	Most expressive	Least expressive	Most expressive
Happy	1171.77 (660.32)	972.96 (610.39)	1433.25 (826.29)	988.49 (619.81)
Surprised	1095.68 (616.59)	813.25 (519.25)	1515.83 (838.83)	931.77 (645.71)
Sad	946.32 (607.35)	875.48 (512.35)	1078.20 (695.44)	1063.83 (681.78)
Angry	846.49 (521.51)	1033.36 (574.00)	1253.01 (706.61)	1333.79 (745.75)
Disgusted	926.00 (534.30)	831.68 (540.52)	1290.03 (791.26)	1094.84 (619.30)
Neutral	995.24 (621.80)	1151.51 (578.08)	1371.30 (915.00)	1458.14 (679.88)

A 6 x 2 x 2 mixed ANOVA was conducted with "Expressiveness" (least, most), "Expressions Viewed" (happy, surprised, sad, angry, disgusted, neutral) as within-subjects factors and "Group" (retrodicting, classifying) as between-subjects factors.

There was a significant main effect of "Expression Viewed", $F(5,220)=10.05$, $p<0.001$, $\eta_p^2=0.19$. Pairwise comparisons with Bonferroni correction revealed that the average gaze duration at the eyes was longer for reactions to happy expressions, as compared to reactions to sad expressions ($p=0.01$), longer for reactions to angry expressions as compared to sad expressions ($p=0.002$) and longer for reactions to neutral expressions than surprised expressions ($p=0.009$), sad ($p<0.001$), angry ($p=0.013$) and disgusted ($p=0.001$). There was also a significant main effect of "Expressiveness", $F(1,44)=16.99$, $p<0.001$, $\eta_p^2=0.28$ suggesting that perceivers looked at the eyes longer when targets' reactions were less expressive. There was, however, no significant main effect of group, $F(1,44)=2.10$, $p=0.16$, $\eta_p^2=0.05$. There was a significant interaction between "Expression Viewed" and "Expressiveness", $F(5, 220)=13.59$, $p<0.001$ (as shown in Figure 5.14). There was no significant interaction between "Expression Viewed" x "Group", $F(5, 220)= 2.19$, $p=0.056$, $\eta_p^2=0.05$, no significant interaction between "Expressiveness and Group", $F(1, 44)=3.95$, $p=0.05$, $\eta_p^2=0.08$ and no significant interaction between "Expression Viewed", "Expressiveness" and "Group", $F(5, 220)=1.28$, $p=0.27$, $\eta_p^2=0.03$.

5.7.2.1.1.1 Post-hoc analyses for interaction between "Expressions Viewed" and "Expressiveness"

Six independent samples *t*-tests (with Bonferroni correction) showed that perceivers looked longer at the eyes for the least expressive stimuli than the most expressive stimuli for reactions to happy [$t(45)=4.65$, $p<0.001$] and surprised [$t(45)=6.25$, $p<0.001$] expressions. Two separate one-way ANOVAs were conducted for each level of expressiveness. For the least expressive, there was a significant main effect of "Expression Viewed", $F(5,225)=8.25$, $p<0.001, \eta_p^2=0.16$. Pairwise comparisons (with Bonferroni correction) revealed that perceivers looked at the eyes longer when targets were reacting to happy than sad ($p=0.003$) and angry expressions ($p<0.001$). Perceivers also looked longer at the eyes for targets' reactions to surprised expressions than sad ($p=0.001$) and angry ($p=0.002$). For the most expressive, there was also a significant main effect of "Expression Viewed", $F(5,225)=15.68$, $p<0.001$. When targets' reactions were more expressive, perceivers looked at the eyes longer for reactions to angry expressions as compared to happy ($p=0.011$), surprised ($p<0.001$), sad ($p=0.005$) and disgusted ($p=0.001$). Perceivers also looked longer at targets' reactions to neutral expressions than happy ($p=0.001$), surprised ($p<0.001$), sad ($p<0.001$) and disgusted ($p<0.001$). The interaction between "Expression Viewed" and "Expressiveness" can be seen from Figure 5.16.

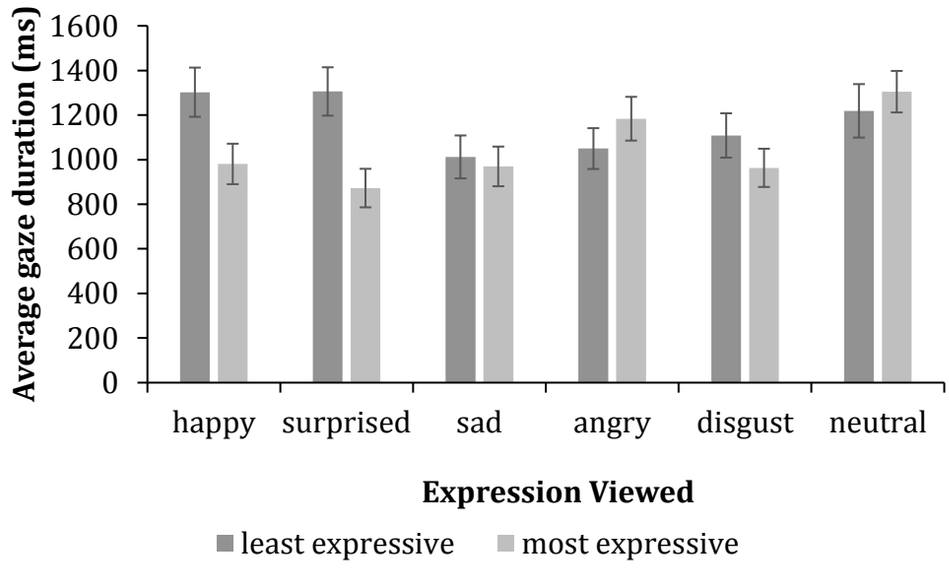


Figure 5.16. The interaction between expressions viewed by targets and expressiveness for the average gaze duration (ms) on the eyes region. Error bars represent the standard error of mean.

5.7.1.2.2 Mouth

Outliers which were identified as values ±2.5 times standard deviation from the median were removed from the analyses (Leys et al., 2013). 12 individual cases out of 552 (2.17%) was removed. The assumption of normality was violated for 5 out of 12 variables for the retrodicting group, *Shapiro-Wilks, p*<0.05. Thus, the data was transformed using square root transformation. Analyses on both transformed and untransformed data revealed the same results, hence, only analyses on transformed data were reported.

The means and standard deviation for each type of reaction, before and after data transformation, is shown in Table 5.16.

Table 5.16 *Mean (and standard deviation) of gaze duration at the mouth for each type of reaction, separated by each perceiver group, after data transformation.*

	Classifying		**Retrodicting**	
	Least expressive	Most expressive	Least expressive	Most expressive
Happy	19.95 (8.04)	25.05 (8.00)	19.73 (8.84)	25.48 (8.75)
Surprised	21.06 (6.92)	23.54 (10.08)	22.14 (10.67)	23.62 (11.21)
Sad	22.97 (10.10)	19.56 (9.84)	19.06 (10.75)	19.42 (10.56)
Angry	24.97 (7.84)	19.56 (9.84)	24.70 (10.78)	21.34 (11.71)
Disgusted	22.95 (9.70)	23.71 (6.02)	20.01 (11.73)	22.25 (7.66)
Neutral	22.07 (6.85)	19.60 (6.96)	20.53 (9.90)	19.60 (8.81)

A 2 x 6 x 2 mixed ANOVA was conducted with "Expressiveness" (least expressive, most expressive), "Expressions Viewed" (happy, surprised, sad, angry, disgusted, neutral) as within-subjects factors and "Group" (retrodicting, classifying) as between-subjects factors. There was a significant main effect of "Expression Viewed", $F(5,195)=4.41$, $p=0.001$, $\eta_p^2=0.10$ whereby perceivers looked longer at the mouth for angry than sad ($p=0.021$) and neutral ($p=0.029$). There was no significant main effect of "Expressiveness", $F(1,39)=0.40$, $p=0.53$, $\eta_p^2=0.01$ and no significant main effect of "Group", $F(1,39)=0.59$, $p=0.45$, $\eta_p^2=0.015$. There was a significant interaction between "Expressiveness" and "Expression Viewed", $F(5,$

195)=5.35, p<0.001, η_p^2=0.012 (as shown in Figure 5.17). However, there was no significant interaction between "Expressiveness" and Group, F(1,39)=1.15, p=0.70, η_p^2=0.004, no significant interaction between "Expressions Viewed" and "Group", F(5, 195)=0.49, p=0.79, η_p^2=0.012 and no significant interaction between "Expressiveness", "Expression Viewed" and "Group", F(5,195)=0.77, p=0.57, η_p^2 =0.019.

5.7.1.2.2.1 Post-hoc for interaction between "Expressiveness" and "Expressions Viewed"

Six paired-samples *t*-tests (with Bonferroni correction) were conducted and revealed that perceivers looked at the mouth significantly longer for most expressive than least expressive reactions, only for happy, t(43)=4.85, p<0.001. Two separate one-way ANOVAs were conducted for each level of expressiveness. For the least expressive reactions, there was a significant main effect of "Expressions Viewed", F(5,205)=2.73, p=0.02, η_p^2=0.06. Perceivers looked longer at the mouth for reactions to angry as compared to reactions to happy (p=0.005) expressions. As for the most expressive reactions, there was a significant main effect of "Expressions Viewed", F(5,210)=7.25, p<0.001, η_p^2=0.15. Perceivers looked at the mouth significantly longer for surprised than neutral (p=0.035) and sad (p=0.015).

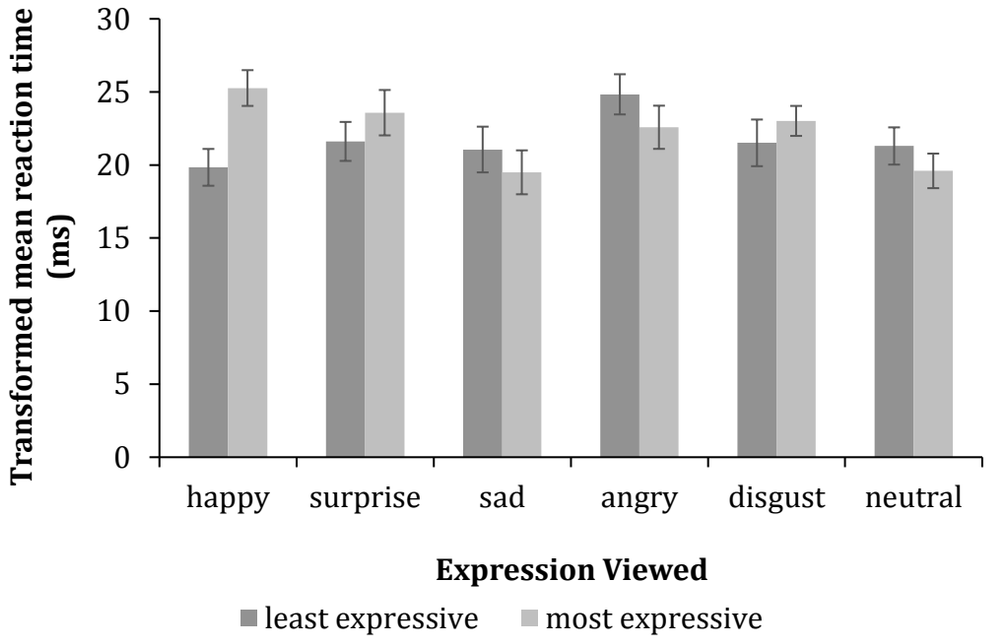


Figure 5.17. The interaction between expressions viewed by targets and expressiveness for the average gaze duration (ms) on the mouth region. Error bars represent the standard error of mean.

5.7.2.2 Differences between classifying and retrodicting perceivers' eye movements across the different time segments

The mean gaze duration across all three time segments for both groups is shown in Table 5.16. Time 1 refers to 0 to 1000ms, Time 2 refers to 1000 to 2000ms while Time 3 refers to 2000 to 3000ms.

Similar to the previous analysis, two separate 3 x 2 ANOVAs were conducted for each area of interest (eyes, mouth), with "Time" (1,2,3) as a within-subjects factor and "Group" (retrodicting, classification) as a between-subjects factor.

5.7.2.2.1 Eyes

No outliers were found according to the Median Absolute Deviation (Leys et al., 2013). The assumption of normality was not violated for all variables for both perceiver groups, Shapiro-Wilks, $p>0.05$. The mean and standard deviation of the gaze duration at the eyes is shown in Table 5.17.

Table 5.17 *Mean (and standard deviation) gaze duration (ms) at the eyes for each perceiver groups, across three time segments.*

	Classifying	**Retrodicting**
T1 (0-1000ms)	254.00 (161.39)	338.81 (193.29)
T2 (1000-2000ms)	355.32 (167.89)	431.86 (222.68)
T3 (2000-3000ms)	319.10 (174.21)	369.64 (206.37)

Mauchly's test of sphericity revealed that the assumption of sphericity was violated for the main effect of "Time", $\chi^2(2)=23.15$, $p<0.001$. There was a significant main effect of "Time", $F(1.48,65.04)=32.20$, $p<0.001$, $\eta_p^2=0.42$. Pairwise comparisons with Bonferroni correction revealed that perceivers looked longer at the eyes during time 2 as compared to time 1 ($p<0.001$) and time 3($p<0.001$). Perceivers also looked longer at the eyes during time 3 as compared to time 1 ($p=0.007$). There was no significant main effect of "Group", $F(1,44)=1.72$, $p=0.20$, $\eta_p^2=0.04$ and there was no significant interaction between "Time" and "Group", $F(1.48, 65.04)=1.09$, $p=0.33$, $\eta_p^2=0.02$.

5.7.2.2.2 Mouth

Outliers were identified and removed according to the Median Absolute Deviation method (Leys et al., 2013). Five out of 138 individual cases (3.6%) were removed. The assumption of normality was violated for 1 variable for the retrodicting group, *Shapiro-Wilks*, $p<0.005$, hence, the data was transformed using square root transformation. As analyses on both untransformed and transformed data yielded the same results, only analyses on transformed data were reported. Table 5.18 shows the mean and standard deviation of gaze duration at the mouth, after data transformation.

Table 5.18 *Mean (and standard deviation) gaze duration (ms) at the eyes for each perceiver groups, across three time segments, after data transformation.*

	Classifying	**Retrodicting**
T1 (0-1000ms)	13.01 (4.12)	12.05 (4.64)
T2 (1000-2000ms)	13.31 (3.66)	12.65 (5.29)
T3 (2000-3000ms)	12.81 (3.44)	12.07 (4.72)

A 3 x 2 mixed ANOVA was conducted with Time (T1, T2, T3) as within-subjects factor and group (retrodicting, classifying) as between-subjects factor. The assumption of sphericity was violated for the main effect of "Time", $\chi^2(2)=12.94$, $p=0.002$. There was no significant main effect of "Time", $F(1.66,69.88)=1.47$, $p=0.26$, $\eta_p^2=0.03$, no significant main effect of "Group", $F(1,42)=0.35$, $p=0.56$, $\eta_p^2=0.008$, and no significant interaction between "Time" and "Group", $F(2,84)=0.16$, $p=0.86$, $\eta_p^2=0.004$.

5.8 DISCUSSION

Similar to Experiment 2:1, perceivers in both groups were slightly above chance level in their responses. Perceivers in the retrodiction group were significantly above chance in inferring which emoticons targets chose while perceivers in the classifying group were significantly above chance in inferring which emoticons best represents targets' expressions. As for reaction time, there was no significant difference between groups, contrary to the findings in Experiment 2:1. Rather, perceivers' response times were different depending on the expression viewed by targets, only for perceivers who were asked to guess which emoticon targets chose ("retrodicting" condition).

In relation to the pattern of eye movements, there was no significant difference between groups in the amount of time they looked at the eyes and mouth. Instead, the pattern of eye movements at the eyes and mouth varied according to the type of reactions perceivers viewed. For instance, for least expressive reactions, perceivers looked at the eyes longer for targets' reactions to happy and surprised expressions, as compared to sad and angry expressions. On the other hand, for most expressive reactions, perceivers' average gaze duration at the eyes was significantly longer for targets' reactions to angry and neutral, as compared to happy, surprised, sad and disgusted expressions. Perceivers also looked at the eyes significantly longer for least expressive stimuli than most expressive stimuli, for reactions to happy and surprised expressions. As for the gaze duration at the mouth, perceivers looked longer at the mouth for targets' reactions to angry than happy expressions, for least expressive reactions. Perceivers looked longer at the mouth for most expressive reactions to surprised

as compared to neutral and sad expressions. Additionally, perceivers also looked at the mouth significantly longer for most expressive than least expressive reactions, only for reactions to happy expressions.

In summary, the findings from this study showed that there was a slight group difference, as demonstrated in the difference in rank response times for different types of expressions. However, there were no overall significant group differences in terms of their discrimination between different expressions, response times and patterns of eye movements.

5.9 General Discussion

In both studies, perceivers were above chance level in inferring a third party's emotional expression from a target's reaction. In so far as targets spontaneously imitated the expressions that they viewed (Wood, 2016), it seems perceivers were able to guess which expressions targets were looking at (or were able to classify the target's expression appropriately). When the inner state was signalled in targets' expressions, this provided clues as to what they were looking at, hence allowing perceivers to guess which emoticons targets chose. These findings support past retrodiction studies in suggesting that people are able to use other minds as a window onto unobservable external events that happened in the world (Pillai et al., 2012; Cassidy et al., 2014; North et al., 2010; Teoh et al., 2017). Classifying perceivers were also systematically accurate in judging which emoticons best represent targets' expressions. There was a group difference in rank accuracy for the different types of expressions but this was only demonstrated in Experiment 2:1. As for reaction times, retrodicting perceivers in Experiment 2:1

took a longer time to respond than classifying perceivers but these findings were not replicated in Experiment 2:2, hence, this finding is not robust.

The eye-tracking results revealed that perceivers' eye movements depended on the different types of target reactions, supporting previous studies which found that the pattern of eye movements varied across different emotions (Eisenbarth and Alpers, 2011) and reactions to different scenarios (Pillai et al., 2012). But, perceivers' patterns of eye movements were not affected by whether they were asked to retrodict or classify the expressions, thus, the pattern of eye movements did not dissociate between the various processes used by the two groups of perceivers

The lack of group differences can be attributed to a few reasons. Firstly, both retrodicting and classifying perceivers may be using the same processing strategy, be it parallel or serial. In relation to the serial route, retrodiction perceivers might have inferred targets' inner states before inferring which emoticons targets chose. Likewise, classifying perceivers might have inferred targets' inner states before deciding which emoticons best represents targets' expressions. Alternatively, both groups might be using a parallel route. Retrodicting perceivers might have retrodicted from targets' behaviour directly to the emoticon chosen by targets, similar to classifying perceivers who inferred which emoticon best represents targets' expressions, bypassing any inference of targets' inner states. Another plausible explanation would be that there might be a difference in the processes used by the two groups but it cannot be detected using eye-tracking and behavioural measures.

Therefore, these two studies have demonstrated that perceivers are capable, albeit merely better than chance, in retrodicting from a behavior to an event that happened in the world, in this case, the unobservable but signalled inner state of a third party. In other words, people are capable of using other minds as a window onto events that happened in the world, provided that the event causes an inner state that is signalled in targets' expressions. But, these expressions are natural and spontaneous expressions. In real life, we often encounter simulated and disguised expressions. It is not known if perceivers will be able to retrodict accurately if people choose to simulate or disguise their expressions. Perhaps, perceivers who are asked to retrodict will perform better than perceivers who are asked to classify expressions, thus, disentangling the various processes involved. Further studies in this thesis will explore people's ability to use other minds as a lens onto events that happened in the world, from simulated expressions.

CHAPTER 6

Can individuals guess if others are simulating their expressions?

EXPERIMENT 3

6.1 INTRODUCTION

The ability to infer other minds is a vital part of our social communication. There is an overwhelming amount of research to suggest that people can infer some things about other minds (Baron-Cohen et al., 1996, 1997) and to use other minds as a window to external events that happened in the world (Pillai et al., 2012, 2014; Cassidy et al., 2014, 2015). These studies rely on the assumption that our inner states are signalled in our facial expressions. However, the degree to which one can perceive inner states in others (and therefore the degree to which one can use other minds as a window onto the world) is limited in that people are able to control their expressions and even dissimulate (DePaulo et al., 2003), a capability that helps to regulate one's social life (Vrij et al., 2008). For example, people may lie or deceive another person to protect them from being hurt or offended. Apart from verbally lying, people can manipulate their expressions in three ways. Firstly, an expression can be neutralized by displaying a neutral expression despite feeling a certain emotion. Secondly, one can simulate an emotional expression while feeling neutral and lastly, one can mask their emotions by displaying an emotional expression which does not correspond to the emotion that they are feeling. For

simplicity's sake, all types of manipulated expressions will be referred to as simulated expressions. Therefore, emotional deception gives rise to another challenge within this field of research such that we do not know whether people are able to 'see through' others' simulated expressions and hence perceive the true underlying state and indeed the events in the world that provoked the inner state.

Despite one's best efforts to manipulate their facial expressions, Ekman (2004) suggested that deception may be identified from nonverbal cues that arise from both cognitive and emotional sources. Cognitive sources are associated with increased cognitive load (Vrij et al., 2001) whereas emotional sources refer to nonverbal leakage of affective experiences associated with lying (such as fear and guilt) which are incongruent with the lie that the deceiver is trying to present (Ekman, 2004; Ekman and Friesen, 1969, 1974). These nonverbal leakage cues manifest in two forms, subtle expressions and micro-expressions. Subtle expressions refer to fragments of suppressed affect displayed by groups of facial muscles which are difficult to control voluntarily (Ekman and Friesen, 1975) while micro-expressions are full muscular expressions of affect that only occur for about $1/25^{th}$ of a second (Ekman and Friesen, 1969; Frank and Ekman, 1997). If people are sensitive to the emotional leakage that manifests as subtle or micro-expressions, then people would be able to see through others' simulated minds. Although these nonverbal cues point to the suggestion that people can utilise these cues to infer simulated states, only a few studies have investigated whether people can identify others' emotional deception and the findings are inconclusive.

In a study by Ekman and O'Sullivan (1991), participants were asked to watch a film showing amputations and burns, eliciting strong and unpleasant

emotions. While watching, they were asked to conceal their negative feelings and convince others that they were watching a pleasant film. There was an overall accuracy of 86% in guessing whether participants were lying or telling the truth. Nevertheless, this study only investigated whether people were capable of detecting emotional deception from concealed negative emotions. It is not known if people are equally capable at detecting emotional deception when others mask their positive emotions. To date, only one other study had investigated whether people are capable of judging the veracity of reactions to a positive and negative scenario (Porter and ten Brinke, 2008). The authors found that observers were capable of identifying if an expression was neutralized, simulated or masked, only for happy and disgusted expressions. However, in this study, participants were told what emotions the actors had to display and participants had to judge the veracity of the expressions. In real life, people are not normally explicitly told what emotions have to be displayed by others. Furthermore, this study also suffers from the problem that actors were instructed to display specified expressions. Although the actors were told to display a neutral expression, simulate a different expression or to mask their expressions, it is not known if the actors were really doing what they were asked to do. As such, participants' judgements regarding the veracity of the expression cannot be evaluated against an objective criterion.

The present study aims to investigate if people are able to see through others' emotional deception by using a 'retrodiction' task. Target participants will be looking at positive and negative pictures from the International Affective Picture System (IAPS) (Lang et al., 1997) and will have to display an expression stated by a cue word which is either congruent or incongruent with the content of the picture. Although it can be assumed that one is displaying a genuine expression

when they are responding to a picture which is congruent with the cue word and disguised when they are responding to a picture which is incongruent with the cue word, it is not known if the emotional content of the picture is embedded in targets' inner state. Thus, it is not clear whether or not targets are displaying what they actually feel (congruent expressions) or concealing what they feel (incongruent expressions). As such, asking perceivers to guess whether targets' expressions are genuine or disguised would pose the problem of an objective answer.

For instance, one's true inner state might be positive while they are looking at a negative picture (for example, a dirty toilet). Thus, they might not necessarily be disguising their expression if they were asked to display an opposing expression (for example, a positive expression) but merely displaying what they were genuinely feeling. In such cases, it would be inaccurate to view targets' expressions as disguised expressions. Likewise, one might be feeling neutral while simulating a positive expression upon viewing a positive picture (for example, an image of kittens). In this case, targets' expressions are not genuine expressions but a simulation of positive expressions. As there is no way of knowing whether the emotional content of the picture is embedded in targets' inner states, we cannot know for certain whether targets were displaying a genuine or disguised expression.

This study, however, bypasses the problem by asking perceivers whether targets were responding to a congruent or incongruent cue word-picture pairing. According to the 'retrodictive mindreading' procedure (Gallese and Goldman, 1998), upon seeing targets' expressions, perceivers will first infer targets' inner state as signalled in their expressions. If perceivers identified a different inner

state other than the cued expression, then perceivers would judge that the expression was an incongruent expression. Likewise, if perceivers were not sensitive to a different inner state, then perceivers would judge that the expression was a congruent expression. Subsequently, perceivers make a judgement of whether targets were reacting to a congruent or incongruent cue word-picture pairing.

Alternatively, perceivers might be able to perform this task without making any inferences about targets' inner states. Rather, they might infer whether targets' expressions were congruent or incongruent, prior to judging whether targets were reacting to a congruent or incongruent cue word-picture pairing. The former explanation involves the inference of targets' inner states, while the latter does not involve this inference. For expressions where targets' inner states are subtle and not obvious in their expressions, they would mistakenly infer that targets' cued expression reflects their actual inner states and infer that the expression is a reaction to a congruent cue word-picture pairing. Thus, it is imperative for perceivers to be sensitive to targets' inner states for them to make an accurate retrodictive judgement about whether targets are responding to a congruent or incongruent cue word-picture pairing.

Hence, it is assumed that perceivers will first infer targets' inner states prior to inferring whether targets were simulating a congruent or incongruent expression. Then, perceivers will infer if targets were reacting to a congruent or incongruent cue word-picture pairing. It is hypothesized that perceivers will successfully discriminate between incongruent cue word-picture reactions and congruent cue word-picture reactions.

6.2 METHODS

This study was separated into two sections: (1) Generation of Target Video Clips and (2) the Perceiver Phase. The study was approved by the Ethics Committee, School of Psychology, University of Nottingham.

6.2.1 Generation of Target Video Clips

The aim of this part of the study was to generate natural, subtle, brief and dynamic reactions. Targets were shown photographs from the International Affective Picture System (IAPS) in order to elicit a reaction. In the first condition, targets were told that they will be participating in a memory test to avoid any suspicion of being recorded (spontaneous). In the second condition, targets were told that they will still be required to participate in a memory test but this time, their expressions will be recorded. They were asked to display an expression stated by the cue word which was either congruent or incongruent with the emotion elicited by the IAPS picture (simulated). Then, these videos were given to a different group of participants (perceivers) who had to determine whether targets' expressions were congruent or incongruent cue word-picture reactions. This method allows an objective point of reference in the sense that the researcher knows whether targets were responding to a congruent or incongruent pairing between the cue word and the picture that targets were looking at. Hence, perceivers' judgements can be evaluated against an objective criterion.

6.2.1.1 Target participants

Twelve participants (6 females, 6 males, *Mean age* = 23.83 years, SD= 2.59 years) were recruited as target participants. They were recruited from the undergraduate and postgraduate student population at the University of

Nottingham. Target participants were paid 2 pounds as an inconvenience allowance. All participants provided written consent for their videos to be used.

6.2.1.2 Materials and Apparatus

Filming of targets' expressions took place in a spacious experimental room within the School of Psychology. They were seated with their backs facing a white background and as soon as the recording began, the researcher left the experimental room. Targets' expressions were covertly filmed using the integral webcam of a Lenovo ThinkPad X240 laptop. The laptop was positioned approximately 60 cm from the targets' seat in such a way that the laptop webcam recorded the targets' face, neck and shoulders against a white background.

The stimuli used were obtained from the International Affective Picture System (IAPS) (Lang et al., 1997), consisting of 1183 pictures, together with standardized ratings of valence, arousal and dominance. Out of these 1183 pictures, 30 pictures (15 positive and 15 negative) were selected as stimuli for this study. These pictures were separated into 3 sets of stimuli. 10 pictures (5 positive and 5 negative) were used in each condition to elicit spontaneous expressions, simulated congruent and incongruent expressions. Due to ethical reasons, extremely negative pictures (including snakes and spiders) were excluded. Both positive and negative pictures were matched in terms of arousal. Examples of negative pictures included drunk driving, malnourished child, bruised leg, vomit, dirty toilet, fire, car accident, crying child, flood and cigarettes. Examples of positive pictures included a seal, kittens, bunnies, puppies, babies, family, children, sunset, ocean, beach, fireworks, skiing and water rafting. The experiment was displayed on PsychoPy v 1.83.01 (Peirce, 2007).

6.2.1.3 Experimental and Filming Procedure

The purpose of the first phase of the study was to elicit spontaneous and natural target reactions to the pictures. As a cover activity, targets were told that they would be taking part in a memory test in which they would need to recall the picture that they saw. A cue word 'READY' was displayed for 3000 ms. Then, the picture was displayed for 5000 ms. In the second phase, targets were also told that they would need to remember the pictures for a memory test. But this time, it was made known to them that their expressions were being recorded and they had to display an expression stated by the cue word, regardless of what the picture made them feel. The cue words were either congruent or incongruent with the pictures. For instance, in the 'congruent' condition, targets saw the cue word 'POSITIVE' preceding a positive picture. In the 'incongruent' condition, targets saw the cue word 'NEGATIVE' preceding a positive picture. As with the first phase of the study, the cue word was displayed for 3000 ms while the picture was displayed for 5000 ms. Target participants had to view 30 different pictures in total (5 positive and 5 negative in each condition). The pictures used to elicit each type of reaction (spontaneous, simulated congruent, simulated incongruent) were counterbalanced across participants. After the experiment, target participants were debriefed as to the true purpose of the experiment. They were asked to complete a video release consent form if they agreed for their videos to be used, But, if they did not provide consent for their videos to be used, their videos and personal information would be destroyed immediately.

6.2.2 Editing and Selection of video clips

Targets' expressions were segmented into 5-second video clips, each corresponding to the type of picture that they were looking at. These videos were

edited using Windows Movie Maker 2.6. The video frames were 384 pixels in width and 288 pixels in height. The audio component was omitted. Since there were 12 targets, a total of 360 video clips were created. Figure 3.1 shows an example of each type of reaction.

[Figure 6.1 was removed to maintain confidentiality]

Figure 6.1. Examples of targets' reactions: a) spontaneous reaction to negative picture, b) spontaneous reaction to positive picture, c) simulated reaction to a negative cue word-negative picture pairing (congruent negative), d) simulated reaction to a positive cue word-positive picture (congruent positive) pairing, e) simulated reaction to a negative cue word-positive picture pairing (incongruent positive), and f) simulated reaction to a positive cue word-negative picture pairing (incongruent negative).

Two-hundred-and-forty video clips (120 simulated reactions to congruent cue word-picture pairing, 120 simulated reactions to incongruent cue word-picture paring) were selected for this study; spontaneous reactions were not presented to

perceivers in this case, to avoid overcomplicating the design and overloading them with trials.

6.2.3. Perceiver Phase

6.2.3.1 Participants

34 perceivers (29 females, 5 males, *Mean age* = 20.55 years, *SD* = 3.71 years) aged between 18 to 30 years old were recruited from the undergraduate and postgraduate student population at the University of Nottingham. As an inconvenience allowance, participants were given 3 pounds (or awarded 1 credit point if they were first year psychology undergraduate students).

6.2.3.2 Materials and Apparatus

In this study, 240 videos of targets' reactions were presented to perceivers. These expressions were targets' reactions to either an incongruent or congruent cue word-picture pairing. These videos were displayed at 768 pixels in width and 574 pixels in height on PsychoPy v 1.83.01 (Peirce, 2007). One-hundred-and-twenty videos were simulated reactions to an incongruent cue word-picture pairing (60 reactions to positive cue word-negative picture pairing; 60 reactions to negative cue word-positive picture pairing) while the other 120 videos were simulated reactions to a congruent cue word-picture pairing (60 reactions to positive cue word-positive picture pairing; 60 reactions to negative cue word-negative picture pairing). Of 240 videos, 120 videos were of female targets and 120 videos were of male targets. The experiment was displayed on PsychoPy v 1.83.01 (Peirce, 2007). In order to avoid fatigue, the stimuli were separated into two equal blocks, each containing 120 stimuli (60 reactions to congruent cue word-picture pairing; 60 reactions to incongruent cue word-picture pairing).

6.2.3.3 *Procedure*

Before the experiment began, perceivers were seated approximately 60 cm from the screen. Perceivers were told that they will be looking at facial expressions which were reactions to a positive or negative picture. They were also told that targets had to display an expression stated by a cue word, which was either congruent or incongruent with the picture. Perceivers were then asked to indicate if targets were responding to a congruent or incongruent cue word-picture pairing, by selecting the 'CONGRUENT' or 'INCONGRUENT' option. After selecting their response, they had to click another box to confirm their answer. Then, they were asked to rate the strength of the expressions from a scale of 1 to 6 (where "1" represents not expressive and "6" represents very expressive). Perceivers were also asked to confirm their response by clicking another confirmation box. They were also told to respond as quickly and as accurately as possible.

6.3 RESULTS

One participant's data was not recorded due to technical issues. Therefore, only data from 33 perceivers were included in the analyses. As some perceivers did not click the confirmation box to assert their response as to whether the expression is congruent or incongruent, their data was not recorded. As such, there were missing data (less than 3%) but this did not systematically vary between experimental conditions.

6.3.1 Were perceivers systematically accurate in guessing whether targets were responding to an incongruent cue word-picture pairing?

In order to determine if there was any bias in perceivers' responses, a confusion matrix was created in order to illustrate the pattern of responses, as shown in Table 6.1.

Table 6.1 *Data count (and percentage) of perceivers' responses. Shaded cells represent accurate responses. Total does not amount to 100% due to missing data.*

	Negative Picture		**Positive Picture**		**Total Responses**
	Congruent	**Incongruent**	**Congruent**	**Incongruent**	
"Congruent" Responses	1340 (67.68%)	1238 (62.52%)	1197 (60.45%)	1294 (65.35%)	5069 (156.00%)
"Incongruent" Responses	594 (30.00%)	701 (35.40%)	752 (37.98%)	647 (32.17%)	2684 (135.55%)
	1934 (97.68%)	1939 (97.92%)	1949 (98.43%)	1931 (97.52%)	

Since there was a bias towards saying that a given expression was a reaction to a congruent cue word-picture pairing, *d*-prime (*d'*) was calculated to identify if perceivers were significantly above chance in guessing whether targets were responding to an incongruent cue word-picture pairing. *d'* is generated from the hit rate and false alarm rate. In this study, 'hit rate' refers to the proportion of occasions when perceivers correctly identified a reaction to an incongruent cue

word-picture pairing as incongruent. The false alarm rate refers to the proportion of occasions when perceivers mistakenly identified a reaction to a congruent cue word-picture pairing as a reaction to an incongruent cue word-picture pairing. Extreme values of 0 are replaced with 0.5/N while extreme values of 1 are replaced with (N-0.5)/N, where N represents the number of signal or noise trials (Macmillan and Kaplan, 1985). Then, *d'* was calculated by subtracting the z-score of the false alarm rate (*fa*) from the *z*-score of the hit rate (*h*) [*d'*=z(*h*)-z(*fa*)], where function $Z(p)$, $p \in [0,1]$ is the inverse of the cumulative Gaussian distribution. *d'* is, therefore, a measure of the distance between the signal and noise distribution and is an indicator of how well perceivers were able to discriminate between different types of stimuli. Table 6.2 shows the mean hit rate, false alarm rate, *z*-score hit rate, *z*-score false alarm rate and *d'*.

Table 6.2 *Mean values of hit rate (raw), false alarm rate (raw), hit rate (z-score), false alarm rate (z-score) and d'.*

Picture Type	Hit rate (raw)	False alarm rate (raw)	Hit rate (*z*-score)	False alarm rate (*z*-score)	*d'*
Positive	0.32	0.38	-0.47	-0.31	-0.16
Negative	0.35	0.30	-0.37	-0.52	0.15

If perceivers successfully discriminated between different stimuli, there would be a higher hit rate as compared to false alarm rate. Hence, the *d'* value would be significantly above the chance level of 0. If, however, if perceivers did not systematically discriminate between conditions, the hit rate would equal the false alarm rate. This would result in a *d'* value of 0 which is comparable to chance level.

One sample *t*-tests were conducted on *d'* values to identify if perceivers' performance was significantly above chance level. When targets were looking at a negative picture, perceivers were systematically above chance in guessing whether targets were responding to an incongruent cue word-picture pairing, $t(32)=2.22$, $p=0.03$. But, when targets were looking at a positive picture, perceivers were significantly below chance in guessing whether targets were responding to an incongruent cue word-picture pairing, $t(32)=-2.64$, $p=0.01$.

Criterion (*c*) was also calculated by multiplying a factor of -0.5 to the sum of the z-score of hits and false alarms, [$c=-0.5*[z(h)+z(fa)]$] to investigate if there was a significant difference in bias in responding. A paired samples t-test revealed that there was a significant difference in the way perceivers responded to both types of stimuli, $t(32)=-2.33$, $p=0.026$. For reactions to positive pictures, perceivers responded that targets were reacting to incongruent cue word-picture pairing more often for congruent cue word-picture pairing, as compared to reactions to incongruent cue word-picture pairing. On the contrary, for reactions to negative pictures, perceivers responded that targets were reacting to an incongruent cue word-picture pairing more often for incongruent cue word-picture pairing than reactions to congruent cue word-picture pairing.

6.3.2 Were perceivers' response times significantly different for different types of stimuli?

Shapiro-Wilk's test of normality indicated that the assumption of normality was violated for all conditions ($p<0.005$) due to a substantial positive skew. Thus, the data was transformed with log transformation to adjust for normality. After log transformation, the assumption of normality was not violated for all conditions

(p>0.1). Because analyses with untransformed and transformed data yielded the same results, only analyses with transformed data were reported. The mean (and standard deviation) of transformed response times are shown in Table 6.3.

Table 6.3 *Mean (and standard deviation) transformed reaction times.*

	Positive Picture	Negative Picture
Congruent	3.44 (0.15)	3.45 (0.16)
Incongruent	3.48 (0.15)	3.45 (0.16)

A 2 x 2 repeated-measure ANOVA was conducted on transformed reaction times, with 'Congruency' (congruent, incongruent) and "Picture Type" (positive, negative) as within-subjects factors. There was no significant main effect of "Congruency", $F(1,32)=3.38$, $p=0.08$, $\eta_p^2=0.09$. There was no significant main effect of "Picture Type", $F(1,32)=1.46$, $p=0.24$, $\eta_p^2=0.04$. There was no significant interaction between "Congruency" and "Picture Type", $F(1,32)=3.60$, $p=0.07$, $\eta_p^2=0.10$.

6.3.3 Did perceivers' intensity ratings discriminate between different types of stimuli?

Mean (and standard deviation) of intensity ratings is shown in Table 6.4.

Table 6.4 *Mean (and standard deviation) of perceivers' intensity ratings for different types of stimuli.*

	Positive Picture	Negative Picture
Congruent	3.22 (0.54)	3.01 (0.57)
Incongruent	3.14 (0.56)	3.25 (0.57)

A 2 x 2 repeated-measures ANOVA was conducted on intensity ratings, with "Congruency" (congruent, incongruent) and "Picture Type" (positive, negative) as within-subject factors. There was a significant main effect of "Congruency", $F(1,32)=8.77$, $p=0.006$, $\eta_p^2=0.22$ whereby perceivers rated incongruent reactions as more expressive than congruent expressions. There was no significant main effect of "Picture Type", $F(1,32)=3.59$, $p=0.07$, $\eta_p^2=0.10$. There was, however, an interaction between "Congruency" and "Picture Type", $F(1,32)=8.98$, $p=0.005$, $\eta_p^2=0.22$.

6.3.2.1 "Congruency" x "Picture Type" interaction

Four paired samples *t*-tests (with Bonferroni correction) were conducted to further investigate this interaction. When targets were looking at a negative picture, perceivers rated incongruent expressions as more expressive than congruent expressions, $t(32)=4.42$, $p<0.001$. When targets were simulating a congruent expression, perceivers rated targets' reactions to positive pictures as more expressive than their reactions to negative pictures, $t(32)=3.68$, $p=0.001$. Figure 6.2 shows this interaction.

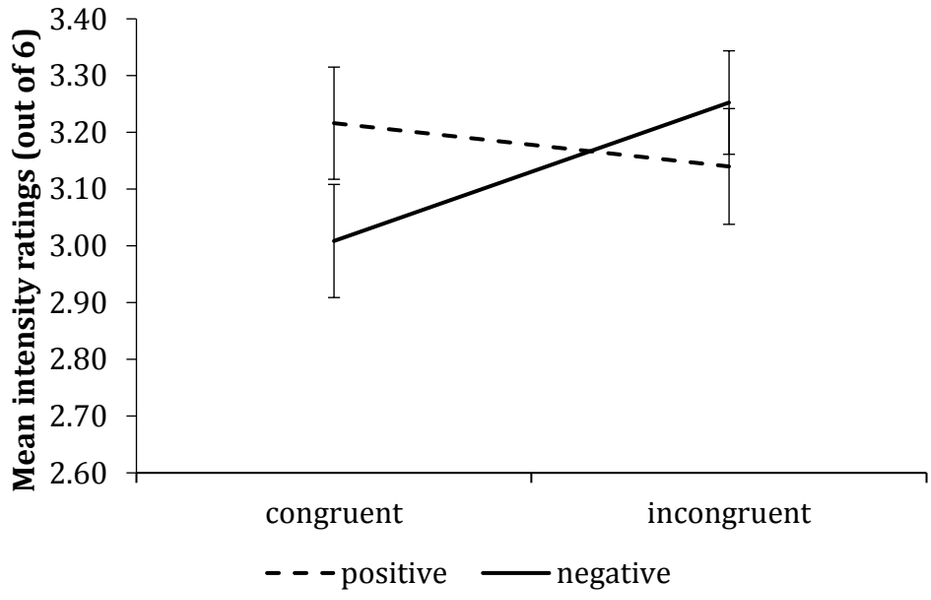


Figure 6.2. Interaction between "Picture Type" viewed by targets and "Congruency" of cue word with the picture. Error bars represent the standard error of mean.

6.4 DISCUSSION

The findings revealed that perceivers performed significantly above chance in guessing whether targets were responding to an incongruent cue word-picture pairing, only for reactions to a negative picture. When targets were looking a negative picture, perceivers were significantly above chance in discriminating between conditions in which targets were asked to display a positive expression (incongruent) and conditions where targets were asked to display a negative expression (congruent). In making their decision as to whether targets were responding to an incongruent cue word-picture pairing, targets must first identify if the expression was congruent or incongruent. Therefore, upon seeing targets' reactions to a negative picture, perceivers were significantly above chance in discriminating between incongruent and congruent expressions. However,

perceivers performed significantly below chance in responding whether targets were responding to an incongruent cue word-picture pairing upon seeing targets' reactions to positive picture. This was because perceivers tended to judge congruent expressions as incongruent more so than incongruent expressions themselves. Thus, when perceivers viewed targets' reactions to a positive picture, they were unable to identify if the expression was congruent or incongruent. Although perceivers had the tendency to judge any given expression as congruent, perceivers judged congruent expressions as congruent more often than incongruent expressions, for reactions to negative pictures. However, for reactions to positive pictures, perceivers judged incongruent expressions as congruent more often than for congruent expressions themselves.

Apart from perceivers' ability to discriminate between different conditions, perceivers' intensity ratings also illustrated that perceivers could discriminate between incongruent and congruent expressions only for reactions to negative pictures. Perceivers rated incongruent expressions to negative pictures significantly stronger than congruent expressions to negative pictures. However, there was no significant difference in intensity ratings between incongruent and congruent expressions to positive pictures.

The discrepancy in findings for both types of reaction could be attributed to the degree of emotional leakage in targets' expressions, which in turn is affected by the extent in which the picture elicited an emotion in targets. In other words, perceivers' ability to infer others simulated minds might be affected by the degree to which the inner state caused by the event is embedded in the target' expression. In Porter and ten Brinke (2008)'s study which utilised the IAPS as stimuli, negative

feelings were more difficult to falsify than positive feelings such as happiness. Furthermore, Warren et al. (2009) also found that while there was an accuracy rate of 64% in detecting lies and truths from emotional stimuli, performance was below chance level in detecting lies and truths from unemotional stimuli. These findings suggest that the ability to detect deception might rely on the amount of emotional leakage present, which in turn is affected by the strength of the stimuli. Likewise, in this study, the negative pictures from the IAPS may have successfully elicited a corresponding negative emotion which is difficult to mask, therefore resulting in emotional leakage. This emotional leakage in the incongruent expressions may have guided perceivers in making accurate judgements about the congruency of the expression. However, positive pictures may have elicited little or no positive emotion in targets, making it easier to simulate a different expression, hence, fooling perceivers. As such, perceivers' ability to infer others simulated minds might be affected by the degree in which the inner state is signalled in targets' expressions. If the inner state caused by the event is not strongly signalled in targets' expressions, perceivers will be easily fooled by targets' simulations.

In summary, these findings revealed that people can identify occasions where others are masking a negative emotion with a positive expression. But, people are fooled by others' expressions when they mask a positive emotion with a negative expression. Although this task involves 'retrodiction' such that perceivers had to guess whether targets were responding to a congruent or incongruent cue word-picture pairing, it is not certain as to whether people are capable of using others' simulated minds as a window onto external events that happened in the world. In other words, people can discriminate between congruent and incongruent reactions to a negative event but it is not known if people can infer

the situational antecedent of an incongruent expression. Thus, the final few studies in the last experimental chapter will explore whether people are capable of retrodicting from spontaneous, simulated congruent and incongruent expressions to an event that caused the expression.

CHAPTER 7

Can individuals retrodict from others' simulated expressions?

EXPERIMENT 4:1

7.1 INTRODUCTION

An effective and successful social exchange is highly dependent on people's ability to predict and infer the inner states of others (Adolphs, 1999). Decades of research have supported the notion that facial expressions are visual print-outs of one's inner states (Baron-Cohen et al., 1997), with findings that people can infer basic (Ekman et al., 1992) and complex emotions (Baron-Cohen et al., 1996, 1997) from posed and exaggerated facial expressions. Furthermore, recent studies have also shown that people are extremely capable of inferring the unobservable situational antecedent of inner states (Pillai et al., 2012, 2014; Cassidy et al., 2014, 2015; Teoh et al., 2017). These studies rely on the assumption that one's true internal states are signalled in our facial expressions and ignores the fact that people are capable of deceiving others. Deception is a common feature of everyday social interaction (Cole, 2001; Jensen et al., 2004; Serota et al., 2010) which serves the purpose of enhancing socially desirable traits and to protect social relationships (DePaulo and Kashy, 1998; DePaulo et al., 1996). As deception forms a fundamental part of social interaction (DePaulo et al., 2003), the ability to infer others' minds also depends on people's ability to tell apart occasions when one is

being honest and deceitful. Many past studies on deception investigated people's ability to identify verbal lies. In classic lie detection studies, observers tended to perform slightly above chance at discriminating between lies and truths (Bond and DePaulo, 2006; Ekman and O'Sullivan, 1991; Vrij et al., 2008). For example, Kraut and Poe (1980) and Vrij et al. (2000) found an accuracy rate of 57% and 56.6% respectively in detecting lies.

Recent studies have found that people can accurately detect deception consciously and unconsciously (ten Brinke et al., 2016). In a study by ten Brinke et al. (2014), people viewed actors telling the truth or lying. Without knowing whether the statements were true or false, participants saw brief subliminal flashes of the actors' faces. When they were asked to classify the statements as truths or lies, they were faster at classifying words as lies after seeing liars' faces and faster at classifying words as honest after seeing the truth tellers' face. In studies that do not inform people whether a friend was lying or telling the truth, people felt less comfortable and more suspicious following a face-to-face interaction in which their friend was lying, although their explicit reports about whether their friend was lying was at chance level (Anderson et al., 2002). Furthermore, when people were kept from consciously deliberating about whether others were lying, their lie detection accuracy increased by 11%, as compared to when they were actively deliberating about whether others were lying (Reinhard et al., 2013). These studies have therefore shown that people are able to detect lies unconsciously.

Apart from lying verbally, people can also deceive others by controlling their facial expressions, such as by simulating an expression of happiness while

actually feeling neutral, by masking one's true emotions by displaying a different expression or by displaying a neutral expression when feeling certain emotion. Surprisingly little attention has been given to whether people can see through others' deceptive emotional displays. In short, we do not know if people are able to 'see through' others' simulated expressions to infer the true underlying inner states of targets and the events which caused those inner states.

Although people can manipulate their expressions, it is possible to infer deception from nonverbal behaviour (Bond et al., 1992; Frank and Ekman, 1997; Vrij et al., 2001) and those who are best at detecting deceit attend more to nonverbal than verbal clues (Ekman and O'Sullivan, 1991). These nonverbal cues arise primarily from both cognitive and affective sources (Ekman, 2001) where the former refers to increased cognitive load (Vrij et al., 2001) and the latter refers to nonverbal leakage of lie-related emotions such as fear and guilt which is incongruent with the state that the deceiver is trying to present (Ekman, 2001; Ekman and Friesen, 1969, 1974). These nonverbal leakages may present themselves as subtle expressions or micro-expressions. For instance, Ekman et al. (1991) found an overall accuracy of 86% in guessing whether others were lying or telling the truth when participants viewed videos of others convincing them that they were watching a pleasant film while they were, in fact, watching a film eliciting strong unpleasant emotions. In another study by Warren et al. (2009), participants were instructed to either give a brief description of hobbies or describe a surgical procedure as though they were watching a Hawaiian beach scene. The accuracy of detecting lies or truths based on responses to emotional stimuli (64%) was significantly higher than unemotional stimuli (36%). As such, these findings support the suggestion that leaked emotions incongruent with the intended

message can provide useful cues to deception. However, the abovementioned studies (Ekman et al., 1991; Warren et al., 2009) only focused on whether people are capable of detecting emotional deception from concealed negative emotions. To date, there has been only one other study which investigated people's ability to judge the veracity of reactions to a positive and negative picture (Porter and ten Brinke, 2008). Observers were capable of identifying deception, only for happy and disgusted expressions. This study, however, suffers from the problem of an objective answer in which it is not known if actors were truly manipulating their expressions in the manner that they were asked to, thus, observers' judgements regarding the veracity of actors' expressions cannot be evaluated against an objective answer. Actors were asked to display a genuine expression, neutralize their expression or disguise their expression. In order for an expression to qualify as a disguised or genuine expression, it is imperative to know one's true inner state. However, as we can never really know for certain as to others' true inner states, it is not certain as to whether they were actually doing what they were asked to do. For instance, upon seeing a positive picture, one might be truly feeling upset. As such, asking perceivers to disguise their expression by putting on a sad expression is not an act of disguise, but a genuine expression.

Even if people were marginally accurate in detecting emotional deception, it is not known if others' simulated minds are transparent to people in such a way that they can use it as a window to infer what caused the true underlying inner state. The present study aims to investigate if people can see through others' emotional deception by employing a "retrodiction" task (Pillai et al., 2012) – Target participants will be looking at positive and negative pictures obtained from the IAPS. In one condition, their expressions will be recorded surreptitiously to obtain

spontaneous reactions. In another condition, they will be told to display an expression stated by a preceding cue word which is either congruent or incongruent with the content of the picture. These natural, congruent and incongruent reactions to positive and negative pictures will be shown to perceivers who have to guess whether targets were looking at a negative picture (Experiment 4:1) or whether targets were looking at a positive picture (Experiment 4:2).

In order to make an accurate response of what targets were looking at, perceivers will have to retrodict from the behaviour of targets to the true underlying inner state of targets (proximal cause) before inferring what picture targets were looking at (distal cause), analogous to a more elaborated and indirect route of processing, as shown in Figure 7.1

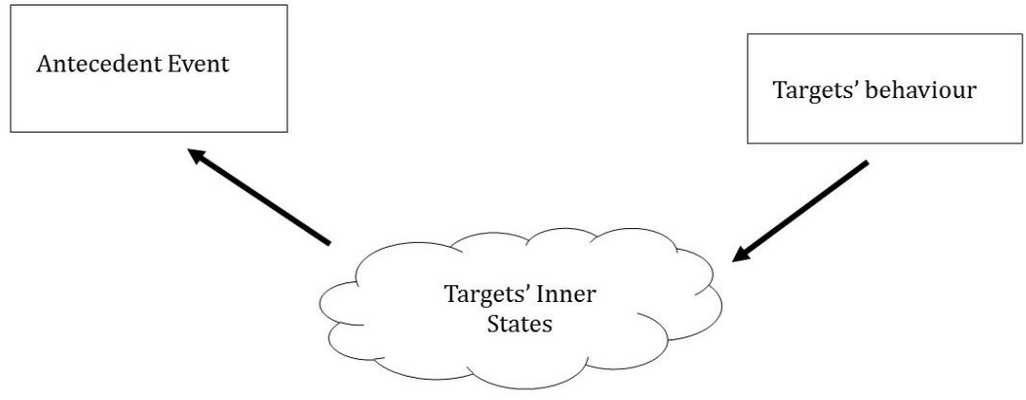


Figure 7.1. Elaborated and indirect route of processing whereby perceivers infer targets' inner states before inferring the antecedent event.

An alternative explanation is that perceivers retrodict directly from targets' behaviour to what targets were looking at, without any inference of targets' inner states, analogous to a simple and direct route of processing (as shown in Figure 7.2.).

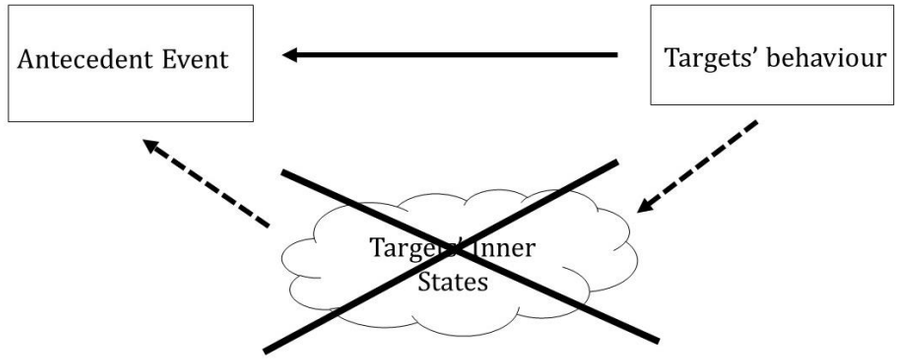


Figure 7.2. Simple and direct route of processing where perceivers retrodict directly from targets' behaviour to the antecedent event, without any inference of inner states.

If retrodicting perceivers engaged in a parallel route of processing, then their responses would be similar to that of perceivers who are merely asked to classify the expression as negative (Experiment 4:1) or positive (Experiment 4:2). We therefore, aimed to compare the responses between retrodicting and classifying perceivers to investigate if retrodicting perceivers genuinely engage in a more elaborated and indirect route of processing. Retrodicting perceivers were asked to explicitly think about targets' inner states when judging whether targets were looking at a negative picture (Experiment 4:1) or a positive picture (Experiment 4:2) whereas classifying perceivers were not explicitly asked to think about targets' inner states when judging if targets' expressions are negative (Experiment 4:1) or positive (Experiment 4:2). In a related topic of predicting guilty and innocent behaviour, Granhag and Hartwig (2008) proposed that 'mindreading' can guide investigative processes to learn about others' strategies, thereby increasing the chances of accurately predicting guilty and innocent behaviour. Thus, explicitly thinking about targets' inner states might improve retrodicting perceivers' performance in inferring what caused targets' simulated

behaviour. As such, we predict that retrodicting perceivers will perform significantly better than classifying perceivers.

In this study, we also recorded the pattern of eye movements. Despite the proposition that people attend more to the eye region in understanding inner states (Argyle, 1978; Baron-Cohen et al., 1997; Hernandez et al., 2009; Janik et al., 1978; Klin et al., 2002), some studies have shown that more attention is given to the mouth region (Cunningham et al., 2004; Kirchner et al., 2011). Other recent studies, however, have indicated that attention to the eyes and mouth vary across different emotions (Eisenbarth and Alpers, 2011) and scenarios (Cassidy et al., 2014; Pillai et al., 2012). Although a vast amount of studies were conducted to investigate people's attention to the eyes and mouth across different emotional states and scenarios, it is not known if their pattern of eye movements vary across different veracity of conditions. As with past research (Cassidy et al., 2014; Eisenbarth and Alpers, 2011; Pillai et al., 2012), average gaze duration was recorded since longer duration could reflect an increase in cognitive processing loads (Goldberg et al., 2000). Therefore, this study aimed to investigate if the pattern of eye movements discriminates between different types of reactions. Additionally, retrodicting perceivers' eye movements will also be compared against classifying perceivers' eye movements, with the aim of disentangling the various processes involved in both groups. A time-course analysis (Freeth et al., 2010) split by 1 second over the duration of stimulus presentation (5 seconds) will also be carried out to investigate if there are any group differences.

Given that findings from Experiment 1 and 2 showed that people are marginally accurate at inferring situational antecedents from spontaneous

expressions, this study, therefore aimed to investigate if perceivers can infer what targets are looking at from spontaneous, congruent and incongruent reactions. We expect that retrodicting perceivers will be able to 'see through' targets' reactions to infer what picture they were really looking at. This study also aimed to investigate if perceivers' behavioural responses and pattern of eye movements discriminate between different types of reactions. Lastly, we also aimed to investigate if retrodicting perceivers engage in a more elaborated route of processing, compared with classifying perceivers. In making their judgements, we expect retrodicting perceivers to engage in an elaborated and indirect form of processing while classifying perceivers engage in a basic and direct form of processing.

7.2 METHODS

This study was approved by the Ethics Committee, School of Psychology, University of Nottingham. This study was separated into two parts: 1) Validation of Stimuli and 2) Perceiver Phase.

7.2.1. Validation of stimuli

The stimuli used was obtained from Experiment 3, that is, targets' spontaneous, congruent and incongruent reactions to positive and negative pictures (IAPS). The aim of this validation stage was to reduce the number of stimuli used in the perceiver phase to minimise discomfort from the eye-tracking headgear. From visual inspection, spontaneous reactions to positive pictures did not elicit any visible facial reactions, suggesting that the emotion elicited by positive pictures were not embedded in targets' expressions. Thus, we aimed to

identify targets who were interpretable from their spontaneous reactions. Thirty judges between 18 to 30 years old (17 females, 13 males, *Mean ag*e=21.63 years old, *SD*=4.26 years) were recruited from the undergraduate and postgraduate student population at the University of Nottingham. They were asked to make a judgement (Yes or No) on whether targets displayed a negative expression. They were also asked to judge the expressiveness of targets' reactions on a scale of 1 to 6 (1=least expressive, 6=most expressive) regardless of whether targets' reactions were negative or not. All judges viewed 120 spontaneous reactions, of which 60 were spontaneous reactions to positive pictures and the remaining 60 were spontaneous reactions to negative pictures. Linear-by-Linear Association tests were conducted for each judge to investigate the relationship between their expressiveness ratings (1, 2, 3, 4, 5, 6) and judges' responses that targets spontaneous reactions were negative (Yes, No). Judges' expressiveness ratings positively correlated with their judgements that the expression is negative, for 23 out of 30 judges (as shown in Table 7.1 and Figure 7.3). Not only that, Linear-by-linear Association tests were also conducted for each judge to investigate the relationship between expressiveness ratings (1, 2, 3, 4, 5, 6) and their accuracy (Accurate, Inaccurate). However, these tests showed that participants' expressiveness ratings did not correlate with accuracy, for all 30 judges (as shown in Table 7.2 and Figure 7.4).

Table 7.1 *Linear-by-linear association values between expressiveness ratings and participants' response that targets expressions were negative, and their associated p-values for all judges.*

Judges	Linear-by-linear association values	*p*-value
1	21.05	$p<0.001$
2	31.61	$p<0.001$
3	34.88	$p<0.001$
4	39.95	$p<0.001$
5	47.33	$p<0.001$
6	34.40	$p<0.001$
7	2.39	$p=0.122$
8	14.54	$p<0.001$
9	5.18	$p=0.02$
10	6.44	$p=0.01$
11	7.65	$p=0.006$
12	39.98	$p<0.001$
13	28.79	$p<0.001$
14	3.33	$p=0.07$
15	0.79	$p=0.38$
16	1.45	$p=0.23$
17	40.32	$p<0.001$

18	12.79	$p<0.001$
19	9.99	$p=0.002$
20	6.14	$p=0.013$
21	14.43	$p<0.001$
22	5.44	$p=0.02$
23	41.81	$p<0.001$
24	1.87	$p=0.17$
25	44.79	$p<0.001$
26	7.80	$p=0.005$
27	2.58	$p=0.11$
28	29.45	$p<0.001$
29	15.27	$p<0.001$
30	17.34	$p<0.001$

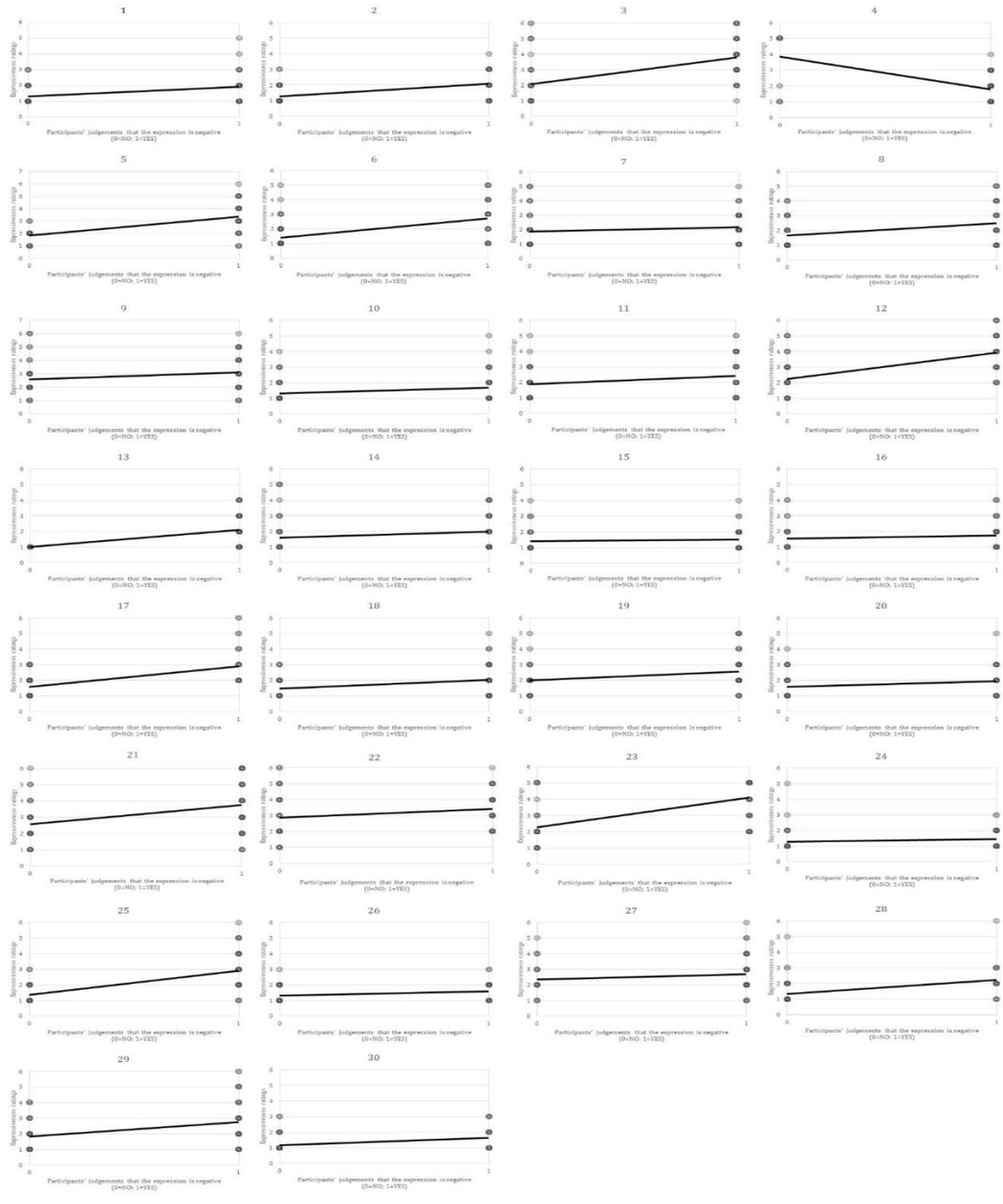


Figure 7.3. The relationship between expressiveness ratings and participants' responses that targets' expressions was negative, for all judges (see Supplementary Materials A for a larger version).

Table 7.2 *Linear-by-linear association values between expressiveness ratings and participants' accurate responses, and their associated p-values for all judges.*

Judges	Linear-by-linear association values	*p*-value
1	0.51	*p*=0.48
2	0.07	*p*=0.79
3	<0.001	*p*=0.99
4	0.97	*p*=0.33
5	1.06	*p*=0.30
6	0.57	*p*=0.45
7	0.20	*p*=0.65
8	2.02	*p*=0.16
9	1.13	*p*=0.29
10	0.87	*p*=0.35
11	0.76	*p*=0.38
12	0.39	*p*=0.53
13	0.24	*p*=0.62
14	2.47	*p*=0.12
15	0.17	*p*=0.68
16	0.67	*p*=0.41
17	0.63	*p*=0.43
18	0.39	*p*=0.53

19	0.13	p=0.72
20	0.81	p=0.37
21	0.25	p=0.62
22	1.15	p=0.28
23	0.01	p=0.91
24	0.73	p=0.39
25	0.21	p=0.65
26	0.03	p=0.85
27	0.03	p=0.87
28	0.03	p=0.87
29	0.64	p=0.42
30	0.01	p=0.91

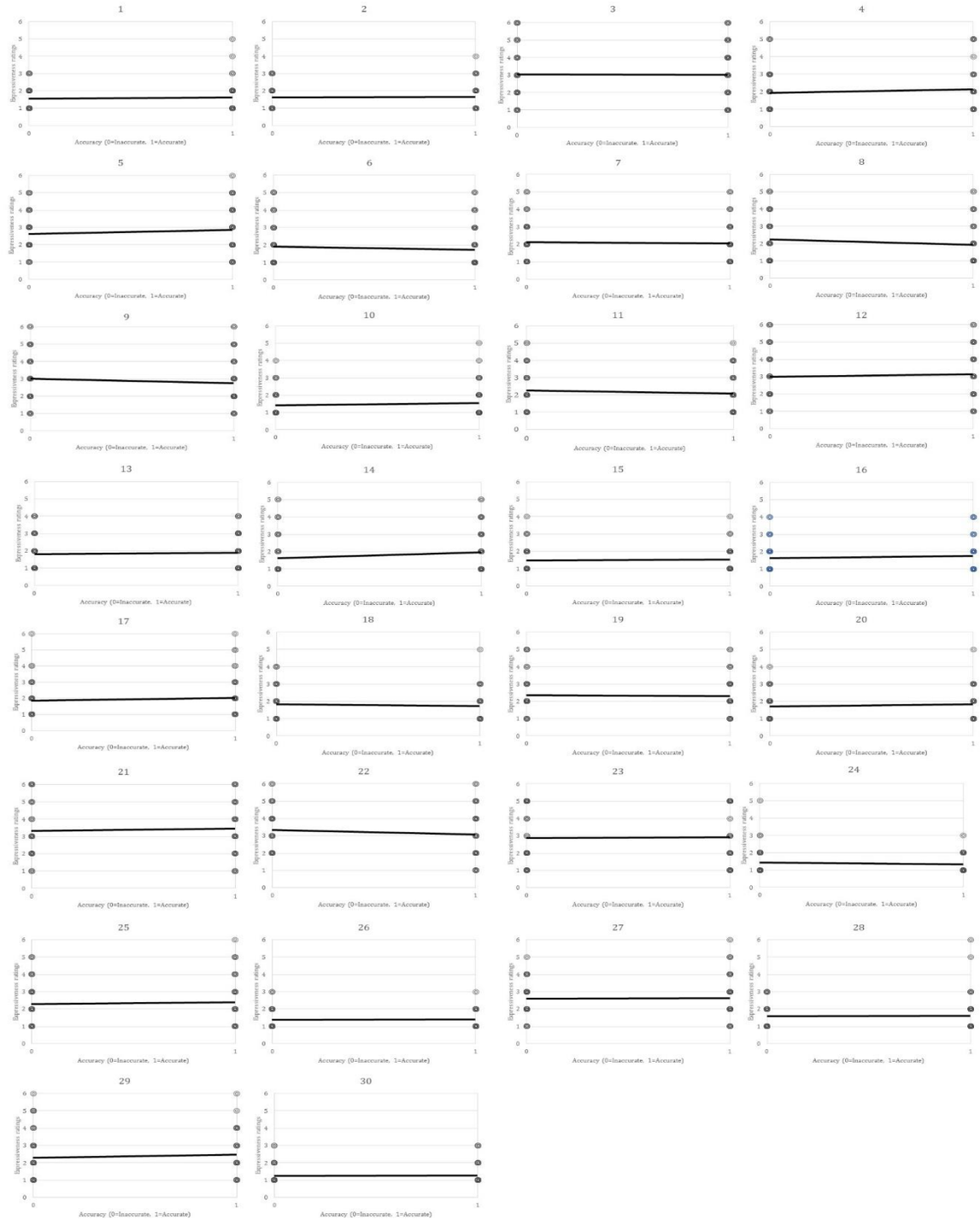


Figure 7.4. The relationship between expressiveness ratings and participants' accurate responses, for all judges (see Supplementary Materials B for a larger version).

Apparently, the stimuli cannot be separated according to their respective expressiveness ratings as they do not correlate with judges detecting whether targets' expressions were negative. As such, targets' individual *d'* values (refer to 3.3.1), the measure of judges' sensitivity in detecting whether targets were reacting to a negative picture, were calculated. In this case, hits refer to occasions where judges accurately identified reactions to negative pictures while false alarms refer to occasions where judges mistakenly identified reactions to positive pictures as reactions to negative pictures. These *d'* values were averaged across all judges to obtain a mean *d'* value for each target. A *d'* value which is significantly above chance level (i.e., zero) indicates that judges could discriminate between targets' spontaneous reactions to negative and positive pictures. According to the targets' mean *d'* values, they were then separated to the "least interpretable targets" (six targets with the lowest *d'* values) and "most interpretable targets" (six targets with the highest *d'* values).

The findings from Experiment 2 suggest that perceivers are only above chance in their responses when targets' spontaneous expressions reflect what they were viewing. As such, if targets' spontaneous expressions do not reflect what they were viewing, targets are not readable and so, there would be no reason to expect that targets' incongruent reactions will be readable. This might potentially give rise to many false positives which will mask the actual findings pertaining to targets who are readable. In such cases, it is also unclear as to whether perceivers' inability to infer targets' incongruent expressions was due to perceivers being fooled by targets' simulations or that the event was not signalled in targets' expressions in the first instance. As for the most interpretable targets, it is unclear as to whether targets' incongruent expressions are readable even if their spontaneous

expressions are readable. On one hand, perceivers might be able to see through targets' incongruent expressions to the actual picture that they are viewing. On another hand, they might be fooled by targets' simulations. Another reason to limit the number of stimuli presented would be to reduce any prolonged discomfort from the eye-tracker. Thus, only the "most interpretable targets" spontaneous, congruent and incongruent reactions were used in the next phase of the study.

7.2.1 Perceiver Phase

7.2.2.1 Design

The design was 'mixed', with three within-subjects factors, "Type of Reactions" (spontaneous, congruent, incongruent), "Cue Word" (positive, negative) which refers to the expression that targets were asked to display and "Picture Valence" (positive, negative) of the pictures that targets were viewing. There was also one between-subjects factor, "Group" (retrodicting, classifying).

7.2.2.2 Participants

40 participants (31 females, 9 males, Mean age=20.13 years, SD=2.63 years) aged between 18 to 30 years old were recruited from the undergraduate and postgraduate student population at the University of Nottingham. Five participants dropped out of the study due to discomfort from the eye tracker head gear while another participant's data was excluded from the analysis due to loss of calibration as most of the fixation points were located out of the screen. The remaining 34 participants (26 females, 8 males, Mean age=20.21 years, SD=2.77 years) were randomly allocated into two groups, retrodicting (12 females, 5 males, Mean age=19.88 years, SD=3.39 years) or classifying (14 females, 3 males, Mean age=20.53 years, 2.03 years).

7.2.2.3 Materials and Apparatus

In this study, 180 videos of "most interpretable" targets' reactions were presented to perceivers. These reactions consist of 60 spontaneous reactions (30 reactions to positive pictures, 30 reactions to negative pictures), 60 congruent reactions (30 reactions to positive cue word-positive picture pairing; 30 reactions to negative cue word-negative picture pairing) and 60 incongruent reactions (30 reactions to negative cue word-positive picture pairing; 30 reactions to positive cue word-negative picture pairing). Of these 180 videos, 120 videos were of female targets (4 unique female targets) while the remaining 60 were of male targets (2 unique male targets). Example stimuli appear in Figure 7.5.

[Figure 7.5 was removed to maintain confidentiality]

Figure 7.5. Example stimuli: a) Spontaneous reaction to a negative picture, b) Spontaneous reaction to a positive picture, c) Congruent reaction to a negative

picture, d) Congruent reaction to a positive picture, e) Incongruent reaction to a negative picture, f) Incongruent reaction to a positive picture.

7.2.2.4 Procedure

Prior to the experiment, perceivers' eye movements were calibrated using a random 9-point calibration system. A fixation dot was presented before each trial to adjust for any drifts in eye movements. The experiment was presented in SR Experiment Builder which was connected to the EyeLink 2.0 eye tracker. All perceivers viewed the expressions for 5 seconds. Classifying perceivers were told to judge if the expression was negative. After viewing the displayed expression, classifying perceivers were asked "Is this expression negative?" to which they had to respond by clicking "Yes" or "No". In contrast, retrodicting perceivers were told that the expressions were either reactions to a negative picture or not. They were also told that these reactions consist of spontaneous, congruent and incongruent reactions. Their task was to judge if targets were looking at a negative picture. After viewing the displayed expression, retrodicting perceivers were asked "Is this person looking at a negative picture?" to which they had to respond by clicking "Yes" or "No". Then, both groups of perceivers had to rate the strength of targets' expressions from a scale of 1 to 6 (1=not at all expressive, 6=very expressive). Targets' expressions were displayed for 5 seconds while the response window was displayed for 30 seconds. Perceivers were told to respond as accurately and as fast as possible. All perceivers viewed 180 reactions, of which 60 were spontaneous reactions, 60 were congruent reactions and 60 were incongruent reactions, presented in a randomised order.

7.3 RESULTS

7.3.1 Behavioural results

7.3.1.1 Were perceivers systematically accurate in their responses?

Table 7.3 *Percentage (and data count) of perceivers' responses for each type of reaction.* Shaded cells represent the correct responses.

		Spontaneous		Congruent		Incongruent	
		Negative picture	Positive picture	Negative picture	Positive picture	Negative picture	Positive picture
Classifying	Negative ("Yes")	67.65% (345)	64.90% (331)	72.75% (371)	13.14% (67)	14.51% (74)	63.53% (324)
	Positive ("No")	32.35% (165)	35.10% (179)	27.25% (139)	86.86% (443)	85.49% (436)	36.47% (186)
		100.00% (510)	100.00% (510)	100.00% (510)	100.00% (510)	100.00% (510)	100.00% (510)
Retrodicting	Negative ("Yes")	69.22% (353)	60.39% (308)	80.78% (412)	13.73% (70)	17.28% (88)	68.82% (351)
	Positive ("No")	30.78% (157)	39.61% (202)	19.22% (98)	86.27% (440)	82.75% (422)	31.18% (159)
		100.00% (510)	100.00% (510)	100.00% (510)	100.00% (510)	100.00% (510)	100.00% (510)

In order to determine if perceivers performed significantly above chance, *d'* was calculated. Hits refers to perceivers accurately judging targets' reactions to negative pictures, while false alarms refer to perceivers mistakenly judging reactions to positive pictures as reactions to negative pictures. The mean *d'* values were compared against a test value of 0 to determine if perceivers' judgments were significantly above chance level. Mean hits, false alarms and *d'* values are shown in Table 7.4.

Table 7.4 *Mean hits, false alarms (raw and normalized) and d' values for both perceiver groups.*

	Classifying			Retrodicting		
	Spontaneous	Congruent	Incongruent	Spontaneous	Congruent	Incongruent
Hits	0.68	0.73	0.18	0.69	0.80	0.20
False Alarms	0.65	0.13	0.60	0.61	0.14	0.66
z(Hits)	0.47	0.61	-0.92	0.50	0.84	-0.84
z(False Alarms)	0.39	-1.13	0.25	0.28	-1.08	0.41
d' value	0.08	1.74	-1.42	0.29	2.06	-1.42

In the retrodicting group, perceivers were significantly above chance level in guessing if targets were looking at a negative picture from spontaneous reactions, $t(16)=4.31$, $p=0.001$ and congruent expressions, $t(16)=9.67$, $p<0.001$. They were systematically inaccurate in their responses for incongruent expressions, $t(16)=-11.80$, $p<0.001$. Classification perceivers on the other hand, were only significantly above chance in judging if the target's expression was negative for congruent expressions, $t(16)=14.83$, $p<0.001$ but not for spontaneous expressions, $t(16)=1.70$, $p=0.11$. Similar to retrodicting perceivers, classifying perceivers were also systematically inaccurate in their responses for incongruent expressions, $t(16)=-12.89$, $p<0.001$.

7.3.1.2 Were perceivers completely fooled by how targets were asked to behave?

The mean proportion of times perceivers judged that targets' expressions were negative was calculated for congruent and incongruent reactions to both negative and positive pictures. The data from both perceiver groups were combined for this analysis. The mean and standard deviation of the proportion of responses appear in Table 7.5.

Table 7.5 *Mean (and standard deviation) for the proportion of times perceivers stated that targets' expressions were negative, for congruent and incongruent reactions.*

	Cue Word	
	Negative	Positive
Congruent	0.76 (0.13)	0.14 (0.08)
Incongruent	0.63 (0.19)	0.19 (0.16)

A 2 x 2 ANOVA was conducted with "Congruency" (congruent, incongruent) and "Cue Word" (positive, negative) as within-subject factors. The results showed a significant main effect of "Congruency", $F(1, 33)=19.94$, $p<0.001$, $\eta_p^2=0.38$, whereby perceivers mentioned that targets' expressions were negative more often in the congruent condition than incongruent condition. There was also a significant main effect of "Cue Word", $F(1, 33)=371.77$, $p<0.001$, $\eta_p^2=0.92$, in which perceivers mentioned that targets' expressions were negative more often when targets were asked to simulate a negative expression than a positive expression. There was also

a significant interaction between "Congruency" and "Cue Word", $F(1, 33)=14.20$, $p=0.001, \eta_p^2=0.30$ as seen in Figure 7.6.

7.3.1.2.1 Post-hoc analyses for interaction between "Congruency" and "Cue Word"

Four paired samples *t*-tests (with Bonferroni correction) were conducted to explain the interaction. Perceivers' judgements that targets' expressions were negative were significantly higher when targets were asked to simulate a negative expression than a positive expression for both congruent [$t(33)=25.55$, $p<0.001$] and incongruent [$t(33)=9.37$, $p<0.001$] conditions. Additionally, perceivers judged that targets' expressions were negative significantly more in the congruent condition than the incongruent condition when targets were asked to simulate a negative expression [$t(33)=4.76$, $p<0.001$], but not when targets were asked to simulate a positive expression [$t(33)=-2.21$, $p=0.04$].

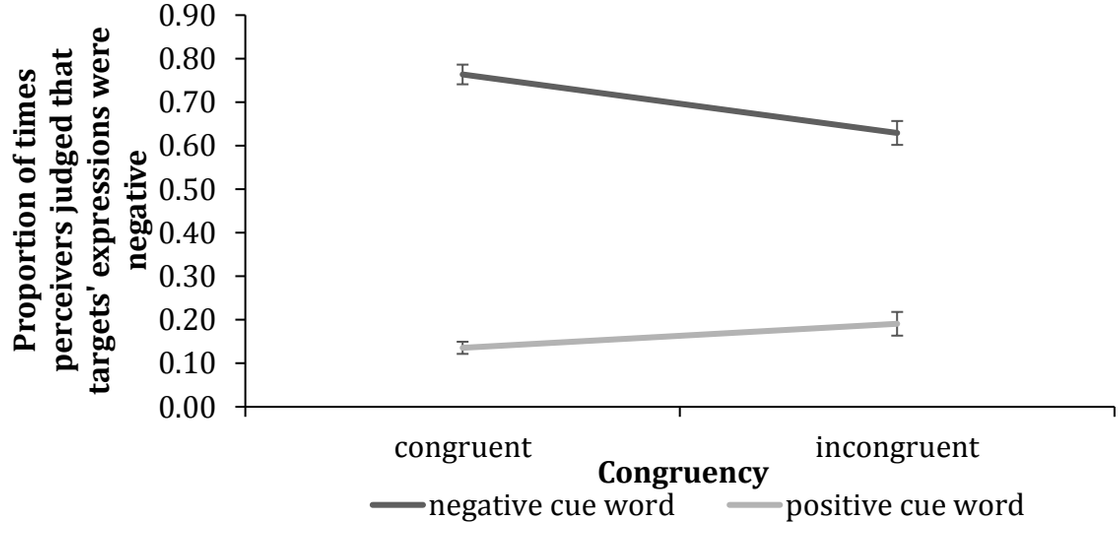


Figure 7.6. Perceivers' judgments that targets' expressions were negative for simulated positive and simulated negative expressions. Error bars represent the standard error of mean.

7.3.1.3 Could perceivers discern whether targets were behaving naturally or responding to a cue?

The mean proportion of times perceivers judged that targets' expressions were negative were compared between targets' spontaneous and congruent reactions to both positive and negative pictures.

The mean and standard deviation of the proportion of responses appears in Table 7.6.

Table 7.6. *Mean (and standard deviation) for the proportion of times perceivers stated that targets' expressions were negative, for spontaneous and congruent reactions.*

	Picture Valence	
	Negative	Positive
Spontaneous	0.68 (0.19)	0.63 (0.19)
Congruent	0.76 (0.19)	0.14 (0.08)

A 2 x 2 ANOVA was conducted with "Type of Reactions" (spontaneous, congruent) and "Picture Valence" (positive, negative). The findings revealed a significant main effect of "Type of Reactions", $F(1,33)=74.39$, $p<0.001$, $\eta_p^2=0.70$, in which perceivers tended to say that spontaneous reactions were negative more often than congruent reactions. There was also a significant main effect of "Picture Valence", $F(1,33)=591.69$, $p<0.001$, $\eta_p^2=0.95$, in which perceivers tended to say that targets' expressions were negative more often when targets were looking at a

negative picture. There was also a significant interaction between "Type of Reactions" and "Picture Valence", $F(1,33)=430.95$, $p<0.001$, $\eta_p^2=0.93$, as shown in Figure 7.7.

7.3.1.3.1 Post-hoc analyses for interaction between "Type of Reactions" and "Picture Valence"

Four paired samples *t*-tests were conducted (with Bonferroni correction). Perceivers judged that targets' expressions were negative more often when targets were looking at a negative picture than a positive picture for both spontaneous [$t(33)=4.38$, $p<0.001$] and congruent [$t(33)=25.55$, $p<0.001$] expressions. When targets were looking at a negative picture, perceivers stated that it was negative more often for congruent than spontaneous, $t(33)=4.02$, $p=0.001$. However, when targets were looking at a positive picture, perceivers stated that the expressions were negative more often for the spontaneous reactions than congruent reactions, $t(33)=14.61$, $p<0.001$.

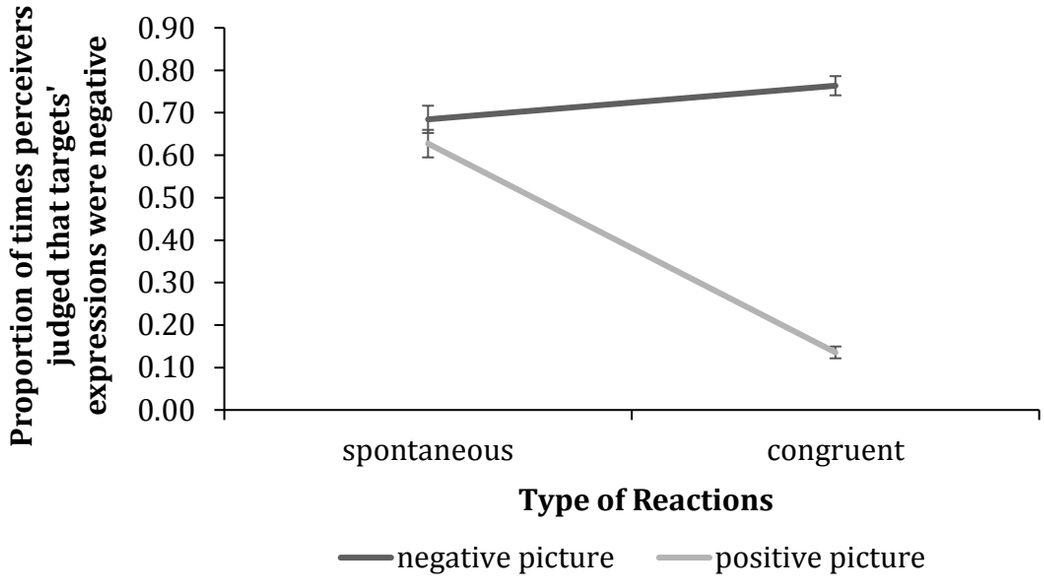


Figure 7.7. Perceivers' judgments that targets' expressions were negative for spontaneous and congruent reactions to positive and negative pictures. Error bars represent the standard error of mean.

7.3.1.4 Were there any significant group differences in perceivers' responses?

A 3 x 2 mixed ANOVA was conducted with "Type of Reactions" (spontaneous, congruent, incongruent) as the within-subjects factor and "Group" (classification, retrodicting) as the between-subjects factor. The assumption of sphericity was violated for the main effect of "Type of Reactions", $\chi^2(2)=38.69$, $p<0.001$. Where the assumption of sphericity was violated, Huynh-Feldt correction was applied. There was a significant main effect of "Type of Reactions", $F(1.22,39.13)=266.25$, $p<0.001, \eta_p^2=0.89$. Pairwise comparisons revealed that perceivers were systematically more accurate for the spontaneous reactions as compared to incongruent reactions ($p<0.001$). Perceivers were also systematically more accurate for congruent reactions as compared to spontaneous ($p<0.001$) and

incongruent reactions (p<0.001). There was also a significant main effect of "Group", $F(1,32)=9.22$, $p=0.005$, $\eta_p^2=0.22$, in which perceivers in the retrodicting group showed higher sensitivity than perceivers in the classification group, as shown in Figure 7.8. There was no significant interaction between "Type of Reactions" and "Group", $F(2,64)=0.65$, $p=0.52$, $\eta_p^2=0.02$.

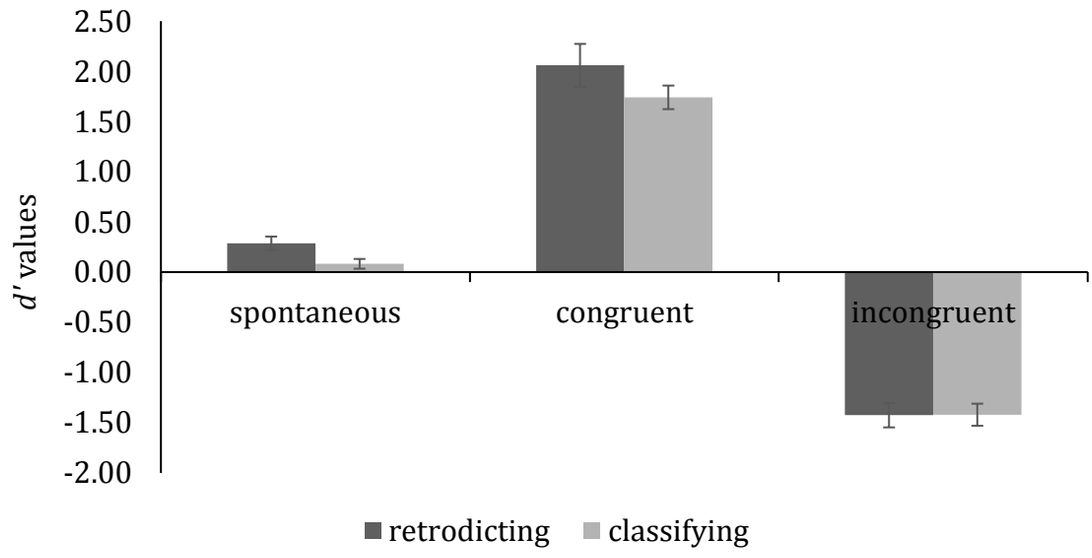


Figure 7.8. Retrodicting and classification perceivers' mean d' values for each type of expression. Error bars represent the standard error of the mean.

7.3.1.5 Were there any significant group differences in perceivers' response times?

The mean and standard deviation (without outliers) is shown in Table 7.7.

Table 7.7 *Mean (and standard deviation) of response times in ms for each type of reaction, for both perceiver groups.*

	Spontaneous		Congruent		Incongruent	
	Negative picture	Positive picture	Negative picture	Positive picture	Negative picture	Positive picture
Classifying	1643.54 (543.58)	1672.78 (495.87)	1520.42 (400.26)	1449.34 (375.79)	1491.92 (397.75)	1433.54 (465.07)
Retrodicting	1386.15 (287.84)	1477.45 (228.96)	1257.52 (286.19)	1251.52 (225.88)	1280.71 (277.60)	1305.84 (284.88)

Outliers ±2.5 standard deviations from the median were identified and removed [6 out of 204 (3%) individual cases] (Leys et al., 2013). The assumption of normality was not violated, *Shapiro-Wilks, p*>0.05. A 3 x 2 x 2 ANOVA was conducted with "Type of Reactions" (spontaneous, congruent, incongruent) and "Picture Valence" (positive, negative) as within-subjects factors and "Group" (retrodicting, classifying) as a between-subjects factor.

There was a significant main effect of "Type of Reactions", $F(2,58)=21.25$, $p<0.001$, $\eta_p^2=0.42$, where perceivers took significantly more time to respond to spontaneous than congruent ($p<0.001$) and incongruent ($p<0.001$) reactions. There was no significant main effect of "Picture Valence", $F(1, 29)=0.36$, $p=0.55$, $\eta_p^2=0.01$, and no significant main effect of "Group", $F(1,29)=2.85$, $p=0.10$, $\eta_p^2=0.09$. There was no significant interaction between "Type of Reactions" and "Group",

$F(2,58)=0.63$, $p=0.54$, $\eta_p^2=0.02$, no significant interaction between "Picture Valence" and "Group", $F(1,29)=0.12$, $p=0.74$, $\eta_p^2=0.004$, no significant interaction between "Type of Reactions" and "Picture Valence", $F(2,58)=1.61$, $p=0.21$, $\eta_p^2=0.05$ and no significant interaction between "Type of Reactions", "Picture Valence", and "Group", $F(2,58)=1.18$, $p=0.31$, $\eta_p^2=0.04$.

7.3.1.6 Did perceivers' expressiveness ratings discriminate the different types of reactions?

The mean and standard deviation of expressiveness ratings appear in Table 7.8.

Table 7.8 *Mean (and standard deviation) of expressiveness ratings (out of 6) for all types of reactions.*

	Spontaneous	Congruent	Incongruent
Negative Picture	2.06 (0.53)	2.61 (0.54)	2.46 (0.49)
Positive Picture	1.90 (0.46)	3.04 (0.49)	2.71 (0.50)

A 3 x 2 ANOVA was conducted on the intensity ratings with "Type of Reactions" (spontaneous, congruent, incongruent) and "Picture Valence" (positive, negative) as within-subjects factors. The assumption of sphericity was violated for the main effect of "Type of Reactions", $\chi^2(2)=19.82$, $p<0.001$, and the interaction between "Type of Reactions" and "Picture Valence", $\chi^2(2)=18.36$, $p<0.001$. There was a significant main effect of "Type of Reactions", $F(1.41, 46.45)=247.75$, $p<0.001$, $\eta_p^2=0.88$. Pairwise comparisons showed that spontaneous reactions were

rated less expressive than congruent reactions, p<0.001 and incongruent reactions, p<0.001. Perceivers also rated congruent reactions as more expressive than incongruent reactions, p<0.001. There was no significant main effect of "Picture Valence", $F(1, 33)=0.25$, $p=0.62$, $\eta_p^2=0.008$. There was a significant interaction between "Type of Reactions" and "Picture Valence", $F(1.43, 47.33)=41.71$, $p<0.001$, $\eta_p^2=0.56$ as shown in Figure 7.9.

7.3.1.6.1 Post-hoc analyses for the interaction between "Type of Reactions" and "Picture Valence"

To explain this interaction, three paired-samples *t*-tests (with Bonferroni correction) were conducted. Reactions to negative pictures were judged to be more expressive than reactions to positive pictures for spontaneous, $t(33)=4.72$, $p<0.001$ and incongruent reactions, $t(33)=5.52$, $p<0.001$. Reactions to positive pictures were judged more expressive as compared to reactions to negative pictures for congruent reactions, $t(33)=6.68$, $p<0.001$. Two one-way ANOVAs were conducted to investigate the effect of "Type of Reactions" on each type of picture viewed by targets. When targets were looking at a negative picture, there was a significant main effect of "Type of Reactions", $F(2,66)=100.09$, $p<0.001$, $\eta_p^2=0.75$, where congruent reactions were judged to be more expressive than spontaneous reactions (p<0.001) and incongruent reactions were judged to be more expressive than spontaneous (p<0.001) and congruent reactions (p=0.004). When targets were looking at a positive picture, there was also a significant main effect of "Type of Reactions", $F(2, 66)=223.82$, $p<0.001$, $\eta_p^2=0.87$, where congruent reactions were judged to be significantly more expressive than spontaneous reactions (p<0.001)

and incongruent reactions (p<0.001). Incongruent reactions were also rated as significantly more expressive than spontaneous reactions (p<0.001).

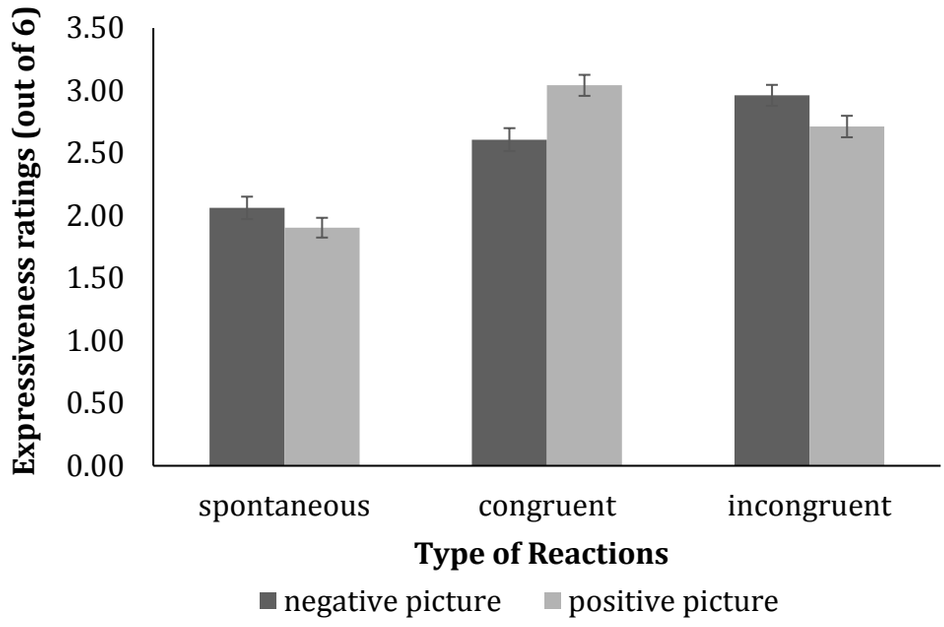


Figure 7.9. Perceivers' intensity ratings for spontaneous, congruent and incongruent reactions to positive and negative pictures.

7.4 DISCUSSION

The findings showed that retrodicting perceivers were significantly above chance level in inferring whether targets were looking at a negative picture, only when targets were reacting spontaneously and when targets were simulating an expression which is congruent with the picture that they were viewing. Classifying perceivers, however, were significantly better than chance level in inferring whether targets' expressions were negative, only when targets were simulating an expression which was congruent with the picture that they were viewing. But, they performed at chance in inferring whether targets' expressions were negative from

spontaneous reactions. Both perceiver groups, nevertheless, were systematically inaccurate when targets were simulating an expression which was incongruent with the picture they were viewing, suggesting that perceivers were fooled by targets' expressions. Retrodicting perceivers were fooled by targets' incongruent reactions to the extent that they could not accurately infer whether targets were looking at a negative picture. Classifying perceivers, too, were fooled by targets' simulations of incongruent reactions in the sense that they were not sensitive to what the picture really made targets feel, indicating that targets' true inner states were not directly perceivable from targets' expressions. If targets' true inner states as caused by the picture viewed were clearly or strongly signalled in targets' expressions, classifying perceivers would not be fooled by targets' simulations. When targets simulated a positive expression while looking at a negative picture, perceivers judged it to be not negative. Likewise, when targets simulated a negative expression while looking at a positive picture, perceivers judged that targets' expressions as negative more often than they judged it to be not negative.

Although it may seem that perceivers were fooled by targets' expressions to the extent that they cannot infer the true inner state of targets and what picture they were looking at, perceivers were not entirely fooled by targets' expressions. Perceivers' frequency of responses in saying that an expression was negative discriminated between conditions in which targets were displaying an incongruent and congruent reaction. When targets were experiencing something positive but cued to display a negative expression, perceivers' judgements that the expression is a reaction to a negative picture reduced significantly, as compared to when targets were genuinely experiencing something negative. On the contrary, when targets were experiencing something negative but cued to display a positive

expression, perceivers' judgements did not significantly differ as compared to when targets were genuinely experiencing something positive. In short, this interaction effect suggests that perceivers could discriminate between congruent and incongruent reactions.

Furthermore, perceivers' responses were also largely accounted for by the way targets were cued rather than the picture that they were looking at. When targets simulated an expression according to a cue, it had a strong effect on perceivers. But without any cue, there was only a small yet systematic effect. When targets were looking at a positive picture, perceivers effectively judged that the expression is not negative when targets were cued. But, when targets were not cued but reacting to a positive picture, targets mostly wore a neutral expression which perceivers might infer as negative. Apart from perceivers' responses, response times, too, differentiated between spontaneous and congruent reactions in such a way that perceivers took more time to respond to spontaneous reactions as compared to cued reactions.

Perceivers' intensity ratings, too, discriminated between reactions. Generally, perceivers rated cued reactions as more expressive than spontaneous reactions. Reactions to a negative picture were judged to be more expressive than reactions to a positive picture when targets were displaying an expression which was incongruent with the picture they were looking at. In contrast, reactions to a positive picture were judged as more expressive as compared to reactions to a negative picture when targets were simulating an expression that was congruent with the picture that they were viewing. Thus, their intensity ratings, too, differentiated between different types of cued expressions. As spontaneous

reactions to negative pictures were judged to be more expressive than spontaneous reactions to positive pictures while the opposite was found for congruent reactions, perceivers' expressiveness ratings thus differentiated between spontaneous and cued reactions.

In judging whether targets were looking at a negative picture, retrodicting perceivers were overall more sensitive than classifying perceivers. This difference in sensitivity suggests different processing strategies were engaged by the different groups. At the very least, for spontaneous reactions, retrodicting perceivers were sensitive to the negative and non-negative cause of the reactions and they were not simply judging that the reaction was negative.

Hence, people cannot infer other minds if others choose to simulate a different mind. Although people may not be completely fooled by others' simulations, this kind of discrimination between congruent and incongruent reactions may not be sufficient to guide an accurate retrodictive judgement to what is the actual cause of the expression. In making retrodictive judgements, perceivers' responses demonstrated that retrodicting perceivers engaged in a different processing strategy than classifying perceivers, where the former possibly engaged in an elaborated and indirect form of processing while the latter possibly engaged in a basic and direct form of processing.

EXPERIMENT 4:2

7.5 INTRODUCTION

This experiment is a replication of Experiment 4:1 but with a small difference. In this study, retrodicting perceivers were asked to judge if targets were looking at a positive picture while classifying perceivers were asked to judge if targets were displaying a positive expression. We also aimed to further disentangle the various processes involved in retrodicting and classifying by comparing accuracy, response times and the pattern of eye movements between both groups.

7.6 METHODS

7.6.1 Perceiver Phase

7.6.1.1 Design

A mixed design was employed with three within-subjects factors, "Type of Reactions" (spontaneous, congruent, incongruent), "Cue Word" (positive, negative) which refers to the expression that targets were asked to display and "Picture Valence" (positive, negative) of the pictures that targets were viewing. There was also a between-subjects factor, "Group" (retrodicting, classifying).

7.6.1.2 Participants

48 participants (31 females, 17 males, *Mean age*=23.08 years, *SD*=3.86 years) aged between 18 to 30 years old were recruited from the undergraduate and postgraduate student population at the University of Nottingham. These participants were randomly allocated to two perceiver groups, retrodicting and classifying. There were 24 participants allocated to the retrodicting group (16

females, 8 males, Mean age=24.04 years, SD=3.07 years) while the remaining 24 participants were allocated to the classifying group (15 females, 9 males, Mean age=22.13 years old, SD=4.44 years).

7.6.1.3 Materials and Apparatus

The stimuli used were identical to Experiment 4:1.

7.6.1.4 Procedure

Perceivers' eye movements were calibrated using a random 9-point calibration system prior to testing. A fixation dot was presented before each trial to adjust for any drifts in eye movements due to the movement of the eye-tracking head gear. The task was presented on SR Experiment Builder which was connected to the EyeLink 2.0 eye tracker. The expressions were displayed for 5 seconds. Classifying perceivers were told to judge if the expression was positive. After viewing the displayed expression, classifying perceivers were asked "Is this expression positive?" to which they had to respond by clicking "Yes" or "No". Retrodicting perceivers were told that the expressions were either reactions to a positive picture or not and that these reactions are either spontaneous, congruent or incongruent with the image displayed to the target. After viewing the displayed expression, retrodicting perceivers were asked "Is this person looking at a positive picture?" to which they had to respond by clicking "Yes" or "No". Then, both groups of perceivers had to rate the strength of targets' expressions on a scale of 1 to 6 (1=not at all expressive, 6=very expressive) regardless of whether it was a positive expression. Targets' expressions were displayed for 5 seconds while the response window was displayed for 30 seconds. Perceivers were told to respond as accurately and quickly as possible. All perceivers viewed 180 reactions, of which

60 were spontaneous, 60 were congruent and 60 were incongruent, presented in randomised order.

7.7 RESULTS

7.7.1 Behavioural results

7.7.1.1 Were perceivers systematically accurate in their responses?

Table 7.9. Percentage (and data count) of perceivers' responses for each type of reaction. Shaded cells represent the correct responses.

		Spontaneous		Congruent		Incongruent	
		Negative picture	Positive picture	Negative picture	Positive picture	Negative picture	Positive picture
Classifying	Negative ("No")	85.00% (612)	80.69% (581)	83.47% (601)	15.14% (109)	17.92% (129)	70.28% (506)
	Positive ("Yes")	15.00% (108)	19.31% (139)	16.53% (119)	84.86% (611)	82.08% (591)	29.72% (214)
	Total responses	100.00% (720)	100.00% (720)	100.00% (720)	100.00% (720)	100.00% (720)	100.00% (720)
Retrodicting	Negative ("No")	83.47% (401)	79.31% (591)	80.14% (577)	14.58% (105)	18.47% (133)	67.92% (489)
	Positive ("Yes")	16.53% (119)	20.69% (149)	19.86% (143)	85.42% (615)	81.53% (587)	32.08% (231)
	Total responses	100.00% (720)	100.00% (720)	100.00% (720)	100.00% (720)	100.00% (720)	100.00% (720)

In order to determine if perceivers performed significantly above chance, *d'* was calculated. Hits refers to perceivers accurately judging targets' reactions to positive pictures, while false alarms refer to perceivers mistakenly judging reactions to negative pictures as reactions to positive pictures. The mean *d'* values (shown in Table 7.10) were compared against a test value of 0 to determine if perceivers' judgments were significantly above chance level.

Table 7.10 *Mean hits, false alarms (raw and normalized) and d' values for both perceiver groups.*

	Classifying			Retrodicting		
	Spontaneous	Congruent	Incongruent	Spontaneous	Congruent	Incongruent
Hits	0.19	0.85	0.30	0.21	0.86	0.32
False Alarms	0.15	0.16	0.82	0.17	0.18	0.81
z(Hits)	-0.88	1.04	-0.52	-0.81	1.08	-0.47
z(False Alarms)	-1.04	-0.99	0.92	-0.95	-0.91	0.88
d' value	0.25	2.20	-1.56	0.21	2.18	-1.48

Perceivers in both groups were significantly above chance in their responses, from spontaneous (Retrodicting: $t(23) = 2.79$, $p=0.010$; Classifying: $t(23)=2.87$, $p=0.009$) and congruent reactions (Retrodicting: $t(23)=17.02$, $p<0.001$; Classifying: $t(23)=17.34$, $p<0.001$). But, both groups of perceivers were significantly below chance in their responses for incongruent reactions (Retrodicting: $t(23)=-13.65$, $p<0.001$; Classifying: $t(23)=-18.53$, $p<0.001$).

7.7.1.2 Were perceivers completely fooled by how targets were asked to behave?

The mean proportion of times perceivers stated that targets' expressions were positive was calculated for congruent and incongruent reactions to both negative and positive pictures. (as shown in Table 7.11) A 2x2 ANOVA was conducted with "Congruency" (congruent, incongruent) and "Cue Word" (positive, negative) as within-subject factors.

Table 7.11 Mean (and standard deviation) of the proportion of times perceivers stated that targets' expressions were positive, for congruent and incongruent reactions.

	Cue Word	
	Negative	Positive
Congruent	0.17 (0.09)	0.85 (0.10)
Incongruent	0.31 (0.10)	0.82 (0.11)

There was a significant main effect of "Congruency", $F(1,44)=54.55$, $p<0.001, \eta_p^2=0.54$, where perceivers made more responses that targets' expressions were positive for incongruent expressions than congruent expressions. There was also a significant main effect of "Cue Word", $F(1,47)=1075.80$, $p<0.001$, $\eta_p^2=0.96$, where perceivers responded that targets' expressions were positive more frequently for targets who were responding to a

positive cue, rather than a negative cue. There was a significant interaction between "Congruency" and "Cue Word", $F(1,47)=124.98$, $p<0.001$, $\eta_p^2=0.73$.

7.7.1.2.1 Post-hoc analyses for interaction between "Congruency" and "Cue Word"

Four paired samples *t*-tests (with Bonferroni correction) were conducted to further investigate this interaction. Perceivers stated that targets' expressions were positive more frequently when targets were responding to a positive cue, as compared to a negative cue for both congruent [$t(47)=34.74$, $p<0.001$] and incongruent [$t(47)=25.51$, $p<0.001$] reactions. When targets were responding to a positive cue, perceivers made more responses that targets' expressions were positive for congruent reactions than incongruent reactions, $t(47)=3.24$, $p=0.002$. But, when targets were responding to a negative cue, perceivers made more responses stating that targets' expressions were positive for incongruent reactions than congruent reactions, $t(47)=14.14$, $p<0.001$. This interaction is illustrated in Figure 7.10.

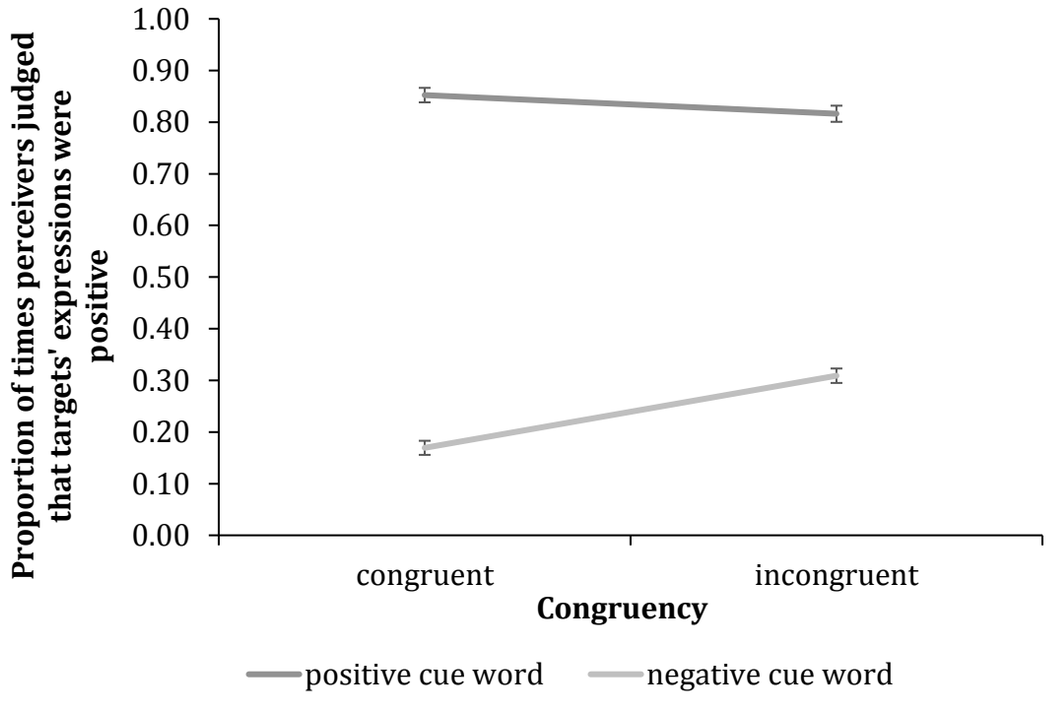


Figure 7.10. Perceivers' responses that targets' expressions were positive for congruent and incongruent reactions to positive and negative pictures. Error bars represent the standard error of mean.

7.7.1.3 Could perceivers discern whether targets were behaving naturally or responding to a cue?

A 2 x 2 repeated-measures ANOVA was conducted on the proportion of times perceivers judged that targets' expressions were positive with "Type of Reactions" (spontaneous, congruent) and "Picture Valence" (positive, negative) as within-subjects factors.

Table 7.12 *Mean (and standard deviation) of the proportion of times perceivers stated that targets' expressions were positive, for spontaneous and congruent reactions.*

	Picture Valence	
	Negative	Positive
Spontaneous	0.16 (0.13)	0.20 (0.13)
Congruent	0.17 (0.09)	0.85 (0.10)

There was a significant main effect of "Type of Reactions", $F(1,42)=535.75$, $p<0.001$, $\eta_p^2=0.92$, where perceivers stated that targets' expressions were positive more frequently for congruent expressions than spontaneous expressions. There was also a significant main effect of "Picture Valence", $F(,1,47)=959.61$, $p<0.001$, $\eta_p^2=0.95$, where perceivers responded that targets' expressions were positive more frequently when targets were reacting to a positive picture as compared to a negative picture. Furthermore, there was a significant interaction between "Type of Reactions" and "Picture Valence", $F(1,47)=772.40$, $p<0.001$, $\eta_p^2=0.94$.

7.7.1.3.1 Post-hoc analyses for interaction between "Type of Reactions" and "Picture Valence"

Four paired-samples *t*-tests were conducted. Perceivers made more responses that targets' expressions were positive for reactions to a positive picture as compared to a negative picture, for spontaneous, $t(47)=3.31$, $p=0.002$ and

congruent, $t(47)=34.74$, $p<0.001$ reactions. Moreover, perceivers judged that targets' expressions were positive more often for congruent reactions than spontaneous reactions when targets were looking at a positive picture, $t(47)=29.36$, $p<0.001$, but not when targets were looking at a negative picture, $t(47)=0.77$, $p=0.45$. This interaction is illustrated in Figure 7.11.

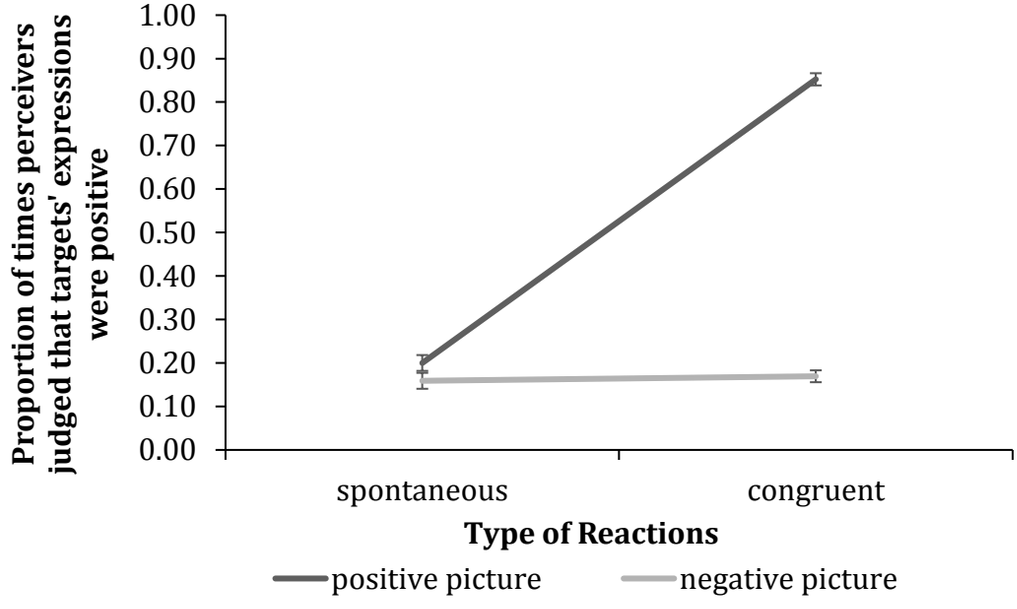


Figure 7.11. Perceivers' responses that targets' expressions were positive for spontaneous and congruent reactions to positive and negative pictures. Error bars represent the standard error of mean.

7.7.1.4 Were there any significant group differences in perceivers' responses?

A 3 x 2 mixed ANOVA was conducted on *d'* with "Type of Reactions" (spontaneous, congruent, incongruent) as a within-subjects factor and "Group" (retrodicting, classifying) as a between-subjects factor. There was a significant main effect of "Type of Reactions", $F(1.58,72.51)=514.77$, $p<0.001$, $\eta_p^2=0.92$ (The

assumption of sphericity was violated, χ^2 (2)=18.08, p<0.001), where perceivers were significantly more sensitive for spontaneous than incongruent expressions (p<0.001). Perceivers were also significantly more sensitive for congruent expressions than spontaneous (p<0.001) and incongruent expressions (p<0.001). There was, however, no significant main effect of "Group", F(1,46)=0.02, p=0.89, η_p^2 <0.001. There was also no significant interaction between "Type of Expression" and "Group", F(2,92)=0.06, p=0.95, η_p^2=0.001.

7.7.1.5 Were there any significant group differences in perceivers' response times?

Outliers which were ±2.5 standard deviations from the median were identified (Leys et al., 2013) and 6 out of 288 (2.1%) individual cases were removed. Since the assumption of normality was violated for one of the variables in the classifying group, (*Shapiro-Wilks, p*<0.05), the data were submitted to square root transformation. The analyses before and after transformation yielded similar results, thus, only analyses after transformation were reported.

The mean and standard deviation (without outliers) appear in Table 7.13.

Table 7.13 *Mean (and standard deviation) of transformed response times in ms for each type of reaction, for both perceiver groups.*

	Spontaneous		Congruent		Incongruent	
	Negative picture	Positive Picture	Negative picture	Positive picture	Negative picture	Positive picture
Retrodicting	37.51 (6.02)	37.75 (5.03)	37.84 (6.38)	36.08 (5.61)	36.97 (5.99)	36.93 (5.85)
Classifying	37.69 (5.52)	38.65 (5.55)	37.58 (6.24)	36.03 (6.19)	36.95 (5.62)	36.71 (5.30)

A 3 x 2 x 2 mixed ANOVA was conducted on reaction times with "Type of Reactions" (spontaneous, congruent, incongruent), "Picture Valence" (positive, negative) as within-subjects factors and "Group" (retrodicting, classifying) as a between-subjects factor. There was a significant main effect of "Type of Reactions", $F(2,86)=6.95$, $p=0.002$, $\eta_p^2=0.14$ whereby perceivers took significantly more time to respond to spontaneous than congruent, $p=0.017$, and incongruent reactions , $p=0.004$. But, there was no significant main effect of "Picture Valence", $F(1,43)=2.75$, $p=0.11$, $\eta_p^2=0.06$, and no significant main effect of "Group" , $F(1,43)=0.003$, $p=0.96$. There was a significant interaction between "Type of Reactions" and "Picture Valence", $F(2,86)=5.54$, $p=0.005$, $\eta_p^2=0.11$. There was, however, no significant interaction between "Picture Valence" and "Group", $F(1,43)=0.27$, $p=0.61$, $\eta_p^2=0.006$, and no significant interaction between "Type of Reactions", "Picture Valence" and "Group", $F(2,86)=0.23$, $p=0.80$, $\eta_p^2=0.005$.

7.7.1.5.1 Post-hoc analyses for interaction between "Type of Reaction" and "Picture Valence".

Three paired-samples *t*-tests (with Bonferroni correction) showed that perceivers took significantly more time to respond to reactions to negative pictures than reactions to positive pictures, only for congruent reactions, $t(45)=3.48$, $p=0.001$. Two one-way ANOVAs were conducted to investigate the effect of "Type of Reactions" for each type of picture that targets viewed. When targets viewed a negative picture, there was no significant main effect of "Type of Reactions", $F(2,90)=0.94$, $p=0.40$, $\eta_p^2=0.02$. But, when targets were looking at a positive picture, there was a significant main effect of "Type of Reactions", $F(2,88)=11.43$, $p<0.001$, $\eta_p^2=0.21$, where perceivers took longer to respond to spontaneous than congruent ($p<0.001$) and incongruent ($p=0.007$) reactions.

7.7.1.6 Did perceivers' expressiveness ratings discriminate between the different types of reactions?

The mean and standard deviation of expressiveness ratings appears in Table 7.14.

Table 7.14 *Mean (and standard deviation) of expressiveness ratings for all types of reactions.*

	Spontaneous	Congruent	Incongruent
Negative Picture	2.14 (0.77)	2.62 (0.82)	3.06 (0.75)
Positive Picture	2.05 (0.76)	3.02 (0.71)	2.79 (0.83)

A 3 x 2 repeated-measures ANOVA was conducted on intensity ratings with "Type of Reactions" (Spontaneous, Congruent, Incongruent) and "Picture Valence" (positive, negative) as within-subjects factors. The assumption of sphericity was violated for the main effect of "Type of Reactions", $\chi^2(2)=77.66$, $p<0.001$ and interaction between "Type of Reactions" and "Picture Valence", $\chi^2(2)=39.50$, $p<0.001$.

There was a significant main effect of "Type of Reactions", $F(1.11, 52.11)=120.14$, $p<0.001$, $\eta_p^2=0.72$. Pairwise comparisons revealed that perceivers rated congruent reactions as more expressive than spontaneous reactions ($p<0.001$). Perceivers also rated incongruent reactions as more expressive than spontaneous ($p<0.001$) and congruent ($p<0001$) reactions. There was no significant main effect of "Picture Valence", $F(1,47)=0.53$, p=0.62, $\eta_p^2=0.01$. But, there was a significant interaction between "Type of Reactions" and "Picture Valence", $F(1.29, 60.54)=22.55$, $p<0.001$, $\eta_p^2=0.32$.

7.7.1.6.1 Post-hoc analyses for interaction between "Type of Reactions" and "Picture Valence"

Three paired-samples *t*-tests were conducted. For congruent expressions, perceivers rated reactions to positive pictures as more expressive than reactions to negative pictures, $t(47)=5.69$, $p<0.001$. But, for incongruent expressions, perceivers rated reactions to negative pictures as more expressive than reactions to positive pictures, $t(47)=3.72$, $p=0.001$. Two separate one-way ANOVAs were conducted for each type of picture. When targets were looking at a negative picture, there was a significant main effect of "Type of Reactions", $F(2,94)=65.13$, $p<0.001$,

η_p^2=0.59. Pairwise comparisons (with Bonferroni correction) showed that incongruent expressions were rated more expressive than spontaneous (p<0.001) and congruent (p<0.001) reactions. When targets were looking at a positive picture, there was a significant main effect of "Type of Reactions", $F(2,94)=89.92$, p<0.001, η_p^2=0.66. Perceivers rated congruent reactions as more expressive than spontaneous (p<0.001) and incongruent (p<0.001) reactions. Incongruent reactions were rated more expressive than spontaneous reactions. This interaction is illustrated in Figure 7.12.

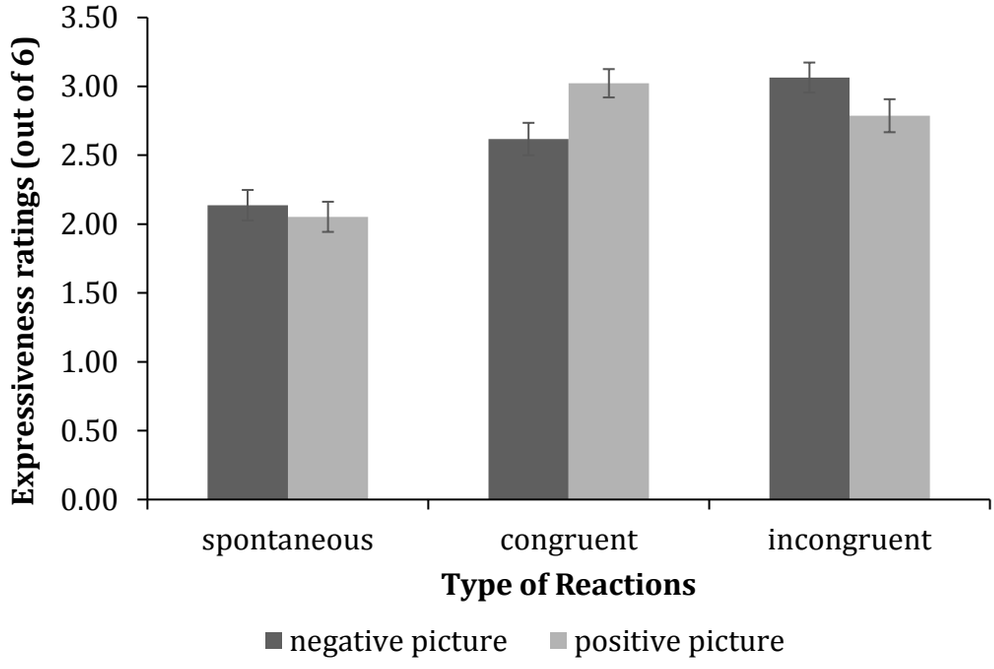


Figure 7.12. Perceivers' expressiveness ratings for spontaneous, congruent and incongruent reactions to positive and negative pictures.

7.7.2 Eye tracking data

Perceivers' eye tracking data were combined across Experiment 4: 1 and Experiment 4: 2. Separate analyses were conducted for each area of interest (Eyes, Mouth).

7.7.2.1 Do perceivers' eye movements discriminate between how targets were behaving?

7.7.2.1.1 Eyes

The mean and standard deviation of average gaze duration at the eyes appears in Table 7.15.

Table 7.15 *Mean (and standard deviation) of average gaze duration, in ms, at the eye region for both studies.*

	Congruent		Incongruent	
	Negative cue	Positive cue	Negative cue	Positive Cue
Experiment 4:1	3335.16 (883.59)	3097.62 (830.53)	3280.62 (868.35)	3086.11 (876.26)
Experiment 4:2	2453.81 (883.56)	3090.61 (1041.84)	3392.90 (866.72)	3117.61 (997.68)

Outliers ±2.5 standard deviations from the median were identified (Leys et al., 2013) and 1 out of 328 individual cases (0.3%) was removed. The assumption of normality was not violated for any of the variables in both experiments, *Shapiro-Wilks, p*>0.05.

A 2 x 2 x 2 mixed ANOVA was conducted with "Congruency" (Congruent, Incongruent), "Cue Word" (positive, negative) as within-subjects factors and "Experiment" (4:1, 4:2) as a between-subjects factor. There was a significant main effect of "Cue Word", $F(1,79)=69.67$, $p<0.001$, $\eta_p^2=0.47$, where average gaze

duration at the eyes was significantly longer when targets were asked to display a negative expression than a positive expression. There was no significant main effect of "Congruency", $F(1,79)=1.82$, $p=0.18$, $\eta_p^2=0.02$ and no significant main effect of "Experiment", $F(1,79)=0.10$, $p=0.75$, $\eta_p^2=0.001$. There was no significant interaction between "Congruency" and "Experiment", $F(1,79)=0.19$, $p=0.67$, $\eta_p^2=0.002$, no significant interaction between "Cue Word" and "Experiment", $F(1,79)=2.59$, $p=0.11$, $\eta_p^2=0.03$, no significant interaction between "Congruency" and "Cue Word", $F(1,79)=3.76$, $p=0.06$, $\eta_p^2=0.05$, and no significant interaction between "Congruency", "Cue Word" and "Experiment", $F(1,79)=0.44$, $p=0.51$, $\eta_p^2=0.006$.

7.7.2.1.2 Mouth

Outliers ±2.5 standard deviations from the median were identified (Leys et al., 2013) and 7 out of 328 individual cases (2.13%) were removed. The assumption of normality was violated for 4 variables in total, hence, data was transformed using a square root procedure. Only analyses with transformed data were reported as analyses before and after transformation yielded same results.

Table 7.16 *Mean (and standard deviation) of transformed average gaze duration, in ms, at the mouth region for both studies.*

	Congruent		Incongruent	
	Negative cue	Positive cue	Negative cue	Positive Cue
Experiment 4:1	36.77 (11.71)	40.54 (10.67)	37.69 (11.28)	41.15 (10.89)
Experiment 4:2	34.50 (10.63)	29.12 (10.72)	35.59 (10.08)	38.81 (11.19)

A 2 x 2 x 2 ANOVA was conducted with "Congruency" (Congruent, Incongruent), "Cue Word" (positive, negative) as within-subjects factors and "Experiment" (4:1, 4:2) as between-subjects factors. There was a significant main effect of "Cue Word", $F(1,77)=88.78$, $p<0.001$, $\eta_p^2=0.54$, where average gaze duration at the mouth was significantly more when targets were displaying a positive expression than a negative expression. There was no significant main effect of "Congruency", $F(1,77)=3.89$, $p=0.05$, $\eta_p^2=0.05$, and no significant main effect of "Experiment", $F(1,77)=0.71$, $p=0.40$, $\eta_p^2=0.009$. There was no significant interaction between "Congruency" and "Experiment", $F(1,77)=0.42$, $p=0.52$, $\eta_p^2=0.005$, no significant interaction between "Cue Word" and "Experiment", $F(1,77)=0.15$, $p=0.70$, $\eta_p^2=0.002$, no significant interaction between "Congruency" and "Cue Word", $F(1,77)=2.73$, $p=0.10$, $\eta_p^2=0.03$ and no significant interaction

between "Congruency", "Cue Word" and "Experiment", $F(1,77)=1.11$, $p=0.30$, $\eta_p^2 =0.01$.

7.7.2.2 Do perceivers' eye movements discriminate between targets behaving naturally and responding to a cue?

7.7.2.2.1 Eyes

Outliers ±2.5 standard deviation from the median were identified (Leys et al., 2013) and 4 out of 328 (1.22%) individual cases were removed. The assumption of normality was not violated, *Shapiro-Wilks, p*>0.05.

Table 7.17 *Mean (and standard deviation) of average gaze duration, in ms, at the eyes region for both studies.*

	Spontaneous		Congruent	
	Negative picture	Positive picture	Negative picture	Positive picture
Experiment 4:1	3306.73 (934.78)	3291.12 (821.77)	3335.16 (883.59)	3041.11 (881.72)
Experiment 4:2	3585.32 (773.02)	3574.64 (784.98)	3453.81 (883.56)	3090.61 (1041.84)

A 2 x 2 x 2 ANOVA was conducted with "Type of Reactions" (spontaneous, congruent), "Picture Valence" (positive, negative) as within-subjects factors and "Experiment" (4:1, 4:2) as between-subjects factor. There was a significant main

effect of "Type of Reactions", $F(1,77)=43.07$, $p<0.001$, $\eta_p^2=0.36$. Average gaze duration at the eyes was significantly longer for spontaneous reactions than congruent reactions. There was also a significant main effect of "Picture Valence", $F(1,77)=51.90$, $p<0.001$, $\eta_p^2=0.40$, where perceivers looked significantly longer at the eyes for reactions to a negative picture than reactions to a positive picture. But, there was no significant main effect of "Experiment", $F(1,77)=0.38$, $p=0.54$, $\eta_p^2=0.005$. There was a significant interaction between "Type of Reactions" and "Experiment", $F(1,77)=5.85$, $p=0.02$, $\eta_p^2=0.07$, and a significant interaction between "Type of Reactions" and "Picture Valence", $F(1,77)=27.10$, $p<0.001$, $\eta_p^2=0.26$. But, there was no significant interaction between "Picture Valence" and "Experiment", $F(1,77)=0.49$, $p=0.48$, $\eta_p^2=0.006$, and no significant interaction between "Type of Reactions", "Picture Valence" and "Experiment", $F(1,77)=3.32$, $p=0.07$, $\eta_p^2=0.04$.

7.7.2.2.1.1 Post-hoc analyses for significant interaction between "Type of Reaction" and "Picture Valence"

Four paired-samples *t*-tests were conducted. Perceivers looked significantly longer at the eyes for reactions to negative pictures as compared to positive pictures for congruent reactions, $t(80)=7.99$, $p<0.001$, but not for spontaneous reactions. Perceivers also looked significantly longer at the eyes for spontaneous reactions as compared to congruent reactions, $t(79)=8.07$, $p<0.001$, when targets were looking at a positive picture but not when targets were looking at a negative picture. This interaction is illustrated in Figure 7.13.

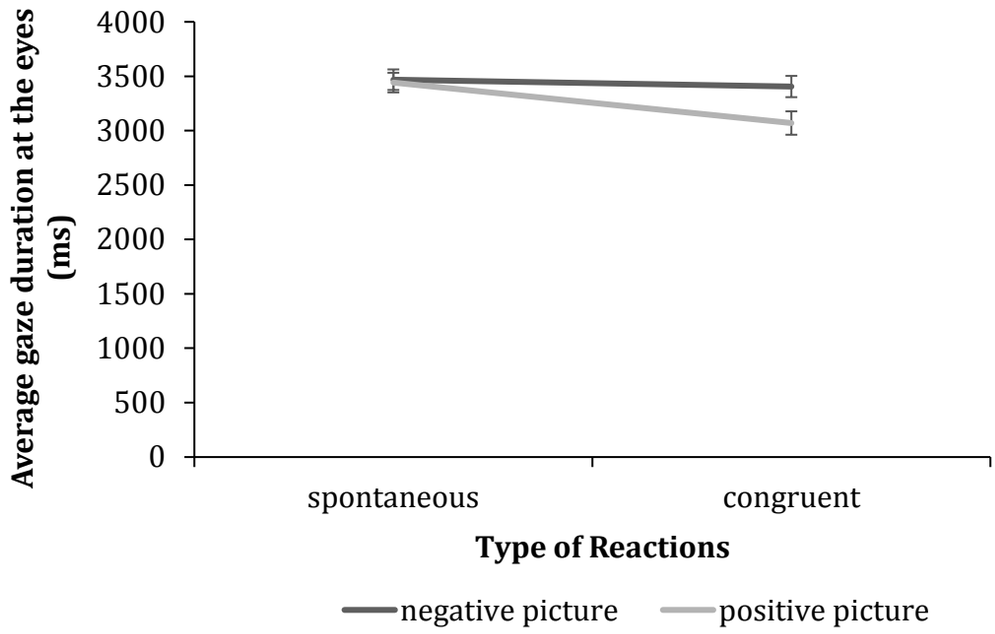


Figure 7.13. Average gaze duration at the eyes when perceivers were looking at congruent and incongruent reactions to positive and negative pictures. Error bars represent the standard error of mean.

7.7.2.2.2 Mouth

Outliers ±2.5 standard deviation from the median were identified (Leys et al., 2013) and 8 out of 328 individual cases (2.43%) were removed. The assumption of normality was violated for 6 variables, thus, the data were transformed using the square root procedure. Since the results were different before and after data transformation, only the analyses with transformed data were reported.

Table 7.18 *Mean (and standard deviation) of transformed average gaze duration, in ms, at the mouth region for both studies.*

	Spontaneous		Congruent	
	Negative picture	Positive picture	Negative picture	Positive picture
Experiment 4:1	37.91 (10.40)	39.25 (9.93)	36.78 (11.71)	40.54 (10.67)
Experiment 4:2	36.57 (9.10)	36.52 (9.86)	34.93 (10.94)	39.68 (11.26)

There was a significant main effect of "Picture Valence", $F(1,77)=49.10$, $p<0.001$, $\eta_p^2=0.39$, where perceivers looked significantly more at the mouth region for reactions to positive pictures than reactions to negative pictures. But, there was no significant main effect of "Type of Reactions", $F(1,77)=1.14$, $p=0.29$, $\eta_p^2=0.15$, and no significant main effect of "Experiment", $F(1,77)=0.22$, $p=0.64$, $\eta_p^2=0.003$. There was a significant interaction between "Type of Reactions" and "Picture Valence", $F(1,77)=29.39$, $p<0.001$, $\eta_p^2=0.28$. But, there was no significant interaction between "Type of Reactions" and "Experiment", $F(1,77)=0.67$, $p=0.42$, $\eta_p^2=0.009$, no significant interaction between "Picture Valence" and "Experiment", $F(1,77)=0.15$, $p=0.70$, $\eta_p^2=0.002$, and no significant interaction between "Type of Reactions", "Picture Valence" and "Experiment", $F(1,77)=3.14$, $p=0.08$, $\eta_p^2=0.04$.

7.7.2.2.2.1 Post-hoc analyses for interaction

between "Type of Reactions" and "Picture Valence"

Four paired-samples *t*-tests were conducted. Perceivers looked significantly longer at the mouth when targets were looking at a positive picture compared to when targets were looking at a negative picture, for congruent reactions, $t(79)=8.40$, $p<0.001$ but not for spontaneous reactions. Average gaze duration at the mouth was significantly longer for spontaneous reactions than congruent reactions when targets were looking at a negative picture, $t(78)=2.79$, $p=0.004$. In contrast, perceivers looked at the mouth significantly longer for congruent reactions than spontaneous reactions when targets were looking at a positive picture, $t(78)=4.36$, $p<0.001$. This interaction is illustrated in Figure 7.14.

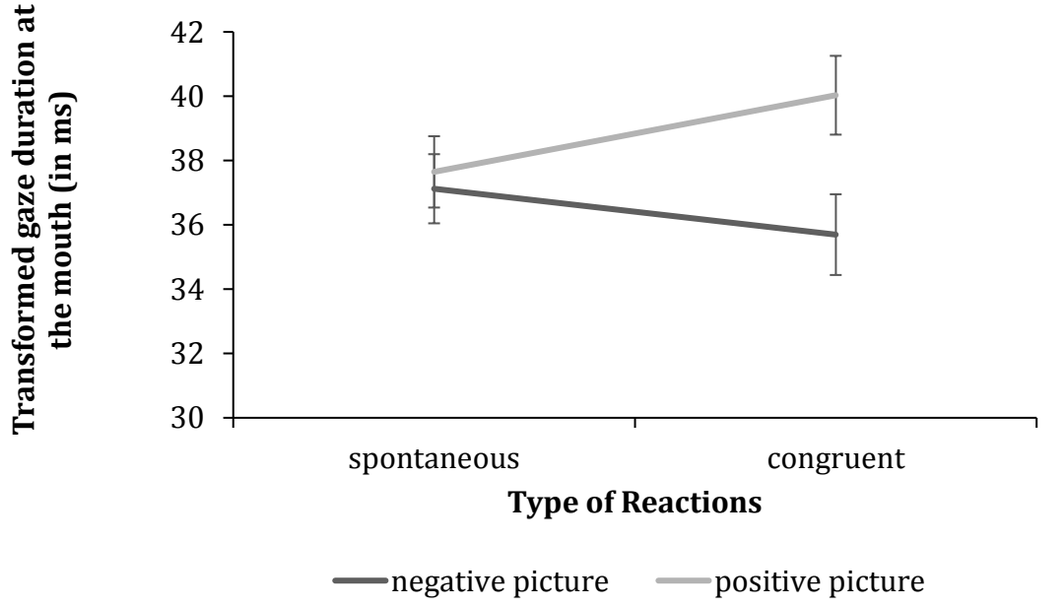


Figure 7.14. Average gaze duration at the mouth when perceivers were looking at spontaneous and congruent reactions to negative and positive pictures. Error bars represent the standard error of mean.

7.7.2.3 Were there any significant group differences in perceivers' eye movements?

Perceivers' eye movements were compared between groups across 5 seconds split by 1.0 second intervals. Post-hoc tests were reported only if there were significant interactions with "Group" as a factor.

7.7.2.3.1 Eyes

Outliers ±2.5 standard deviation from the median were identified (Leys et al., 2013) and 1 out of 410 (0.2%) individual cases were removed. The assumption of normality was violated for 2 variables, and the data were submitted to square root transformation. Only analyses with transformed data were reported.

Table 7.19 Mean (and standard deviation) of transformed average gaze duration, in ms, at the eyes across 5 seconds (T1-T5) for both perceiver groups.

	Experiment 4:1		Experiment 4:2	
	Retrodicting	Classifying	Retrodicting	Classifying
T1	19.31 (4.69)	18.81 (5.11)	21.20 (4.27)	20.12 (4.79)
T2	22.68 (4.81)	21.89 (5.17)	22.08 (4.20)	21.86 (4.04)
T3	22.93 (4.65)	21.83 (4.85)	21.73 (3.97)	22.92 (4.56)
T4	23.08 (5.01)	22.14 (4.74)	23.82 (4.42)	23.51 (4.70)
T5	23.33 (4.53)	22.46 (4.17)	24.19 (3.93)	23.92 (4.46)

A 5 x 2 x 2 mixed ANOVA was conducted with "Time" [T1(0-1000ms), T2(1000-2000ms), T3(2000-3000ms), T4(3000-4000ms), T5(4000-5000ms)] as a

within-subjects factor and "Group" (retrodicting, classifying) and "Experiment" (4:1, 4:2) as between-subjects factors. The assumption of sphericity was violated for the main effect of "Time", $\chi^2(9)=81.24$, $p<0.001$. There was a significant main effect of "Time", $F(3.09, 237.88)=88.27$, $p<0.001$, $\eta_p^2=0.53$, where the average gaze duration at the eyes was significantly lower at T1 than T2 ($p<0.001$), T3 ($p<0.001$), T4 ($p<0.001$) and T5($p<0.001$); T2 as compared to T4 ($p=0.001$) and T5 ($p<0.001$); T3 as compared to T4 ($p<0.001$) and T5 ($p<0.002$); and T4 as compared to T5 ($p=0.031$). There was no significant main effect of "Group", $F(1,77)=0.25$, $p=0.62$, $\eta_p^2=0.003$ and "Experiment", $F(1,77)=0.49$, $p=0.49$, $\eta_p^2=0.006$. There was a significant interaction between "Time" and "Experiment", $F(4, 308)=7.58$, $p<0.001$, $\eta_p^2=0.09$. But, there was no significant interaction between "Time" and "Group", $F(4,308)=1.09$, $p=0.36$, $\eta_p^2=0.01$, and no significant interaction between "Group" and "Experiment", $F(1,77)=0.13$, $p=0.72$, $\eta_p^2=0.002$. There was a significant interaction between "Time", "Group" and "Experiment", $F(4,308)=2.87$, $p=0.023$, $\eta_p^2=0.04$ (as shown in Figure 7.15).

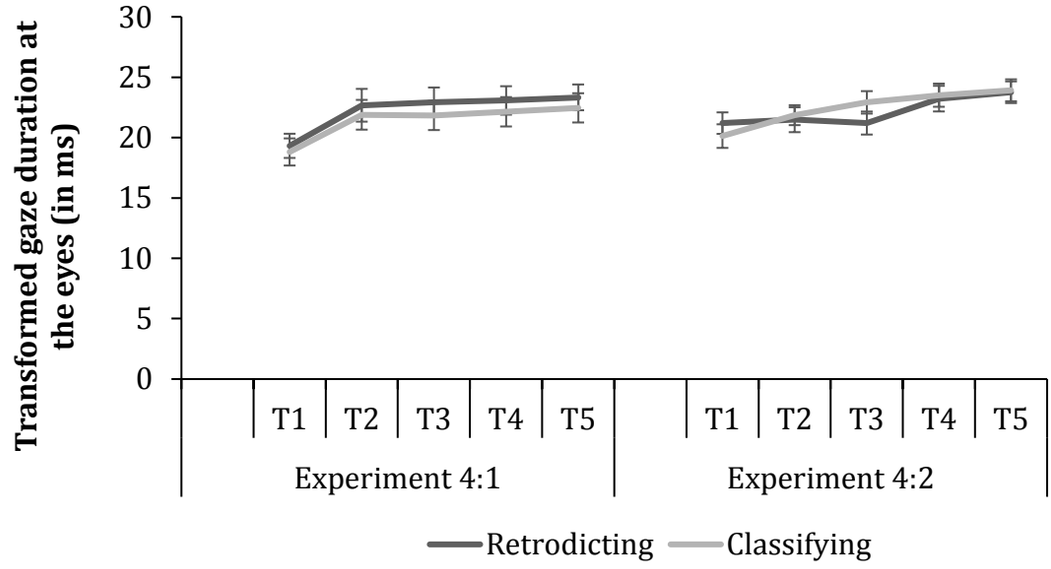


Figure 7.15. Retrodicting and classifying perceivers' average gaze duration at the eyes across 5 seconds for both experiments. Error bars represent the standard error of mean.

7.7.2.3.1.1 Post-hoc analyses for interaction between "Time", "Group" and "Experiment"

Two-way ANOVAs were conducted for each experiment separately. A 5 x 2 mixed ANOVA with "Time" [T1(0-1000ms), T2(1000-2000ms), T3(2000-3000ms), T4(3000-4000ms), T5(4000-5000ms)] as a within-subjects factor and "Group" (retrodicting, classifying) as a between-subjects factor was conducted for both experiments.

In Experiment 4:1, there was a significant main effect of "Time", $F(2.28, 72.90)=57.11$, $p<0.001, \eta_p^2=0.64$ [The assumption of sphericity was violated, $\chi^2(9)=82.59$, $p<0.001$], where average gaze duration at the eyes was significantly less at T1 than T2 ($p<0.001$), T3 ($p<0.001$), T4 ($p<0.001$) and T5($p<0.001$). There

was no significant main effect of "Group", $F(1, 32)=0.28$, $p=0.60$, $\eta_p^2=0.009$, and no significant interaction between "Time" and "Group", $F(4, 128)=0.28$, $p=0.89$, $\eta_p^2=0.009$.

In Experiment 4:2, there was a significant main effect of "Time", $F(3.05, 137.41)=43.62$, $p<0.001$, $\eta_p^2=0.49$ [The assumption of sphericity was violated, $\chi^2(9)=52.05$, $p<0.001$], where average gaze duration at T1 was significantly less than T2 ($p=0.007$), T3 ($p<0.001$), T4 ($p<0.001$) and T5 ($p<0.001$). Average gaze duration was also significantly less at T2 than T4 ($p<0.001$) and T5 ($p<0.001$); T3 than T4 ($p<0.001$) and T5 ($p<0.001$). There was no significant main effect of "Group", $F(1,45)=0.013$, $p=0.91$, $\eta_p^2<0.001$. There was a significant interaction between "Time" and "Group", $F(4,180)=3.90$, $p=0.005$, $\eta_p^2=0.08$. Two separate one-way ANOVAs were conducted for each perceiver group. In the retrodicting group, there was a significant main effect of "Time", $F(2.09, 46.01)=24.55$, $p<0.001$, $\eta_p^2=0.53$ [The assumption of sphericity was violated, $\chi^2(9)=42.71$, $p<0.001$], where average gaze duration was significantly lower at T1 than T4 ($p<0.001$) and T5 ($p<0.001$); T2 than T4 ($p=0.016$) and T5 ($p=0.001$); T3 than T4 ($p=0.003$) and T5 ($p<0.001$). In the classifying group, there was also a significant main effect of "Time", $F(2.55, 58.58)=23.51$, $p<0.001$, $\eta_p^2=0.51$ [The assumption of sphericity was violated, $\chi^2(9)=52.19$, $p<0.001$], where average gaze duration was significantly lower at T1 than T3 ($p<0.001$), T4 ($p<0.001$) and T5 ($p<0.001$); T2 than T5 ($p=0.011$); and T3 than T5 ($p=0.04$). The pattern of ranking of average gaze duration appears in Table 7.20. There were no significant group differences.

Table 7.20 *Pattern of ranking for perceivers' eye movements from highest to lowest.*

Retrodicting	Classifying
T5 ↓ (p>0.05) T4 ↓ (p=0.02) T2 ↓ (p>0.05) T3 ↓ (p>0.05) T1	T5 ↓ (p=0.04) T4 ↓ (p=0.05) T2 ↓ (p>0.05) T3 ↓ (p>0.05) T1

7.7.2.3.2 Mouth

Outliers ±2.5 standard deviation from the median were identified (Leys et al., 2013) and 16 out of 410 (3.99%) individual cases were removed. The assumption of normality was violated for all variables (*Shapiro-Wilks, p*<0.05), and the data were submitted to square root transformation. As analyses before and after data transformation yielded similar findings, only analyses with transformed data was reported.

Table 7.21 *Mean (and standard deviation) of transformed average gaze duration at the mouth across 5 seconds for both perceiver groups.*

	Experiment 4:1		Experiment 4:2	
	Retrodicting	Classifying	Retrodicting	Classifying
T1	15.49 (4.01)	16.20 (4.49)	15.11 (4.12)	16.13 (4.29)
T2	13.22 (5.43)	14.55 (4.93)	12.09 (4.80)	13.37 (4.60)
T3	12.75 (4.87)	14.25 (4.82)	12.55 (4.56)	12.97 (4.41)
T4	11.98 (4.71)	13.79 (4.86)	12.87 (4.90)	12.44 (4.34)
T5	11.22 (4.27)	13.39 (4.99)	12.37 (4.60)	11.35 (4.38)

A 5 x 2 x 2 mixed ANOVA was conducted with "Time" [T1(0-1000ms), T2(1000-2000ms), T3(2000-3000ms), T4(3000-4000ms), T5(4000-5000ms)] as a within-subjects factor, and "Group" (retrodicting, classifying) and "Experiment" (4:1, 4:2) as between-subjects factors. There was a significant main effect of "Time", $F(2.71, 195.35)=86.13$, $p<0.001, \eta_p^2=0.55$ [The assumption of sphericity was violated, $\chi^2(9)=120.42$, $p<0.001$], where average gaze duration was significantly longer at T1 than T2 ($p<0.001$), T3 ($p<0.001$), T4 ($p<0.001$) and T5 ($p<0.001$); T2

than T5(p<0.001); T3 than T4 (p=0.034) and T5 (p<0.001); and T4 than T5 (p<0.001). There was no significant main effect of "Group", $F(1, 72)=0.73$, $p=0.39$, $\eta_p^2=0.01$, and "Experiment", $F(1,72)=1.04$, $p=0.31$, $\eta_p^2=0.014$. There was no significant interaction between "Time" and "Experiment", $F(4,288)=1.50$, $p=0.20$, $\eta_p^2=0.02$, no significant interaction between "Time" and "Group", $F(4,288)=1.72$, $p=0.15$, $\eta_p^2=0.02$, and no significant interaction between "Group" and "Experiment", $F(4,288)=0.35$, $p=0.56$, $\eta_p^2=0.005$. But, there was a significant interaction between "Time", "Experiment" and "Group", $F(4,288)=4.35$, $p=0.002$, $\eta_p^2=0.06$ (as shown in Figure 7.16).

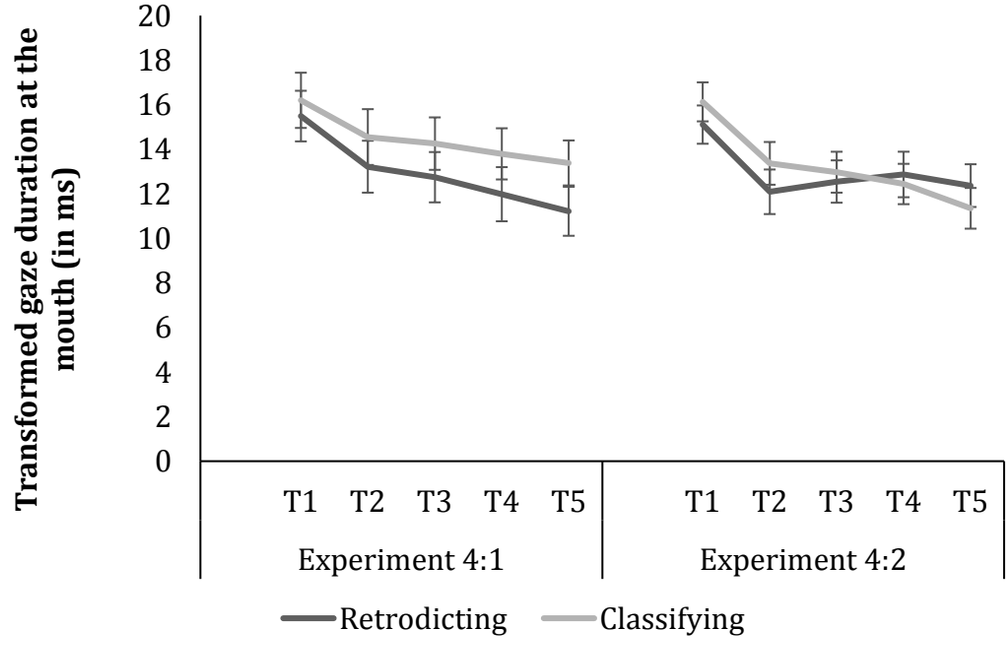


Figure 7.16. Retrodicting and classifying perceivers' average gaze duration at the mouth across 5 seconds for both experiments. Error bars represent the standard error of mean.

7.7.2.3.2.1 Post-hoc analyses for interaction between "Time", "Experiment" and "Group"

Two-way ANOVAs were conducted for each experiment separately. A 5 x 2 mixed ANOVA with "Time" [T1(0-1000ms), T2(1000-2000ms), T3(2000-3000ms), T4(3000-4000ms), T5(4000-5000ms)] as a within-subjects factor and "Group" (retrodicting, classifying) as a between-subjects factor was conducted for both experiments.

In Experiment 4:1, there was a significant main effect of "Time", $F(2.56, 76.78)=38.14$, $p<0.001$, $\eta_p^2=0.56$ [The assumption of sphericity was violated, $\chi^2(9)=59.72$, $p<0.001$], where perceivers looked at the mouth significantly longer at T1 than T2 ($p<0.001$), T3($p<0.001$), T4 ($p<0.001$) and T5 ($p<0.001$); T2 than T5 ($p<0.001$); T3 than T5 ($p<0.001$); T4 than T5 ($p<0.001$). There was no significant main effect of "Group", $F(1,30)=0.75$, $p=0.40$, $\eta_p^2=0.02$. There was no significant interaction between "Time" and "Group", $F(4,120)=0.84$, $p=0.51$, $\eta_p^2=0.03$.

In Experiment 4:2, there was a significant main effect of "Time", $F(2.76, 115.76)=51.63$, $p<0.001$, $\eta_p^2=0.55$ [The assumption of sphericity was violated, $\chi^2(9)=68.12$, $p<0.001$], where perceivers looked at the mouth significantly longer at T1 than T2 ($p<0.001$), T3($p<0.001$), T4($p<0.001$) and T5($p<0.001$); T2 than T5 ($p=0.042$); T3 than T5($p=0.001$); and T4 than T5 ($p<0.001$). There was no significant main effect of "Group", $F(1,42)=0.05$, $p=0.82$, $\eta_p^2=0.001$. There was a significant interaction between "Time" and "Group", $F(4,168)=6.06$, $p<0.001$, $\eta_p^2=0.13$. Separate one-way ANOVAs were conducted for each perceiver group. In the

retrodicting group, there was a significant main effect of "Time", $F(2.71, 56.84)=17.47$, $p<0.001$, $\eta_p^2=0.45$ [The assumption of sphericity was violated, $\chi^2(9)=43.08$, $p<0.001$], where perceivers looked at the mouth significantly less at T1 than T2 ($p<0.001$), T3 ($p=0.001$), T4 ($p=0.001$) and T5 ($p<0.001$). In the classifying group, there was a significant main effect of "Time", $F(2.73, 57.40)=43.45$, $p<0.001$, $\eta_p^2=0.67$ where perceivers looked at the mouth significantly less at T1 than T2 ($p=0.002$), T3($p<0.001$), T4 ($p<0.001$) and T5 ($p<0.001$); T2 than T5 ($p<0.001$); T3 than T5 ($p<0.001$); and T4 than T5 ($p<0.001$). The pattern of ranking of average gaze duration appears in Table 7.23. There were no significant group differences.

Table 7.22 *Pattern of ranking for perceivers' eye movements at the mouth from highest to lowest.*

Retrodicting	Classifying
T1 ↓ (p=0.001) T4 ↓ (p>0.05) T3 ↓ (p>0.05) T5 ↓ (p>0.05) T2	T1 ↓ (p=0.002) T2 ↓ (p>0.05) T3 ↓ (p>0.05) T4 ↓ (p<0.001) T5

7.8 DISCUSSION

The findings showed that retrodicting perceivers were above chance level in their responses when viewing spontaneous and congruent reactions. Retrodicting perceivers were sensitive to whether targets were looking at a positive picture when targets reacted spontaneously and when targets displayed a reaction that is congruent with what they were viewing. Classifying perceivers, too, were above chance in guessing whether targets' expressions were positive when targets were reacting spontaneously and when targets displayed an expression congruent with the picture they viewed. However, when targets displayed an expression which was incongruent with what the picture made them feel, retrodicting perceivers were systematically inaccurate in judging whether targets were looking at a positive picture. Similarly, classifying perceivers were systematically inaccurate in judging if targets' expressions were positive.

At first glance, these findings suggest that perceivers were fooled by targets' reactions. However, even if perceivers cannot retrodict to the actual cause of the expression, they could nevertheless discriminate on some level between different types of reactions, as shown by perceivers' judgments, expressiveness ratings and response times. When targets displayed a positive expression, perceivers judged that targets' expressions were positive more often when targets were indeed looking at a positive picture (congruent), but not when targets were looking at a negative picture (incongruent). When targets were displaying a negative expression, perceivers judged that targets' expressions were positive more frequently when targets were looking at a positive picture (incongruent) as compared to when targets were looking at a negative picture (congruent).

Additionally, perceivers rated reactions to positive pictures as more expressive than reactions to negative pictures, when targets displayed an expression which is congruent with what they were viewing. But, perceivers rated incongruent reactions to negative pictures as more expressive than incongruent reactions to positive pictures. Thus, perceivers' responses and expressiveness ratings discriminated occasions when targets were displaying congruent and incongruent reactions.

Furthermore, perceivers could discriminate when targets were responding naturally and when they were simulating a reaction, from perceivers' responses, response times, expressiveness ratings and pattern of eye movements. Firstly, perceivers judged that an expression was positive more often for reactions to a positive picture as compared to a negative picture, on occasions when targets were reacting spontaneously as well as when they were cued. Perceivers also judged that targets' expressions were positive more frequently for congruent than spontaneous when targets were looking at a positive picture, but not when targets were looking at a negative picture. Moreover, perceivers took a longer time to respond to spontaneous reactions than cued reactions. Perceivers also rated spontaneous reactions as less expressive compared to cued reactions. Perceivers rated spontaneous reactions to a negative picture as less expressive than incongruent reactions to a negative picture. Meanwhile, perceivers rated spontaneous reactions to a positive picture as less expressive than congruent reactions to a positive picture. As for the pattern of eye movements, perceivers looked longer at the eyes when targets were reacting to a negative picture than a positive picture for congruent reactions. Perceivers also looked at the eyes significantly longer for spontaneous reactions than congruent reactions, when

targets were looking at a positive picture but not when targets were looking at a negative picture. In relation to gaze duration at the mouth region, perceivers looked longer at the mouth for reactions to positive pictures as compared to reactions to negative pictures only for congruent reactions, but not for spontaneous reactions. Perceivers looked longer at the mouth for spontaneous reactions than congruent reactions when targets were viewing negative pictures but perceivers looked longer at the mouth for congruent reactions than spontaneous reactions to positive pictures. Therefore, perceivers' responses, reaction times, expressiveness ratings and pattern of eye movements discriminated between occasions when targets were behaving naturally and when they were cued.

Retrodicting perceivers were significantly above chance level in retrodicting from spontaneous and congruent reactions to whether targets were looking at a positive picture. Similarly, classifying perceivers were significantly accurate in judging whether targets' expressions were positive. But, retrodicting perceivers' responses did not significantly differ from classifying perceivers' responses. Although the pattern of eye movements across 5 seconds at the mouth was different for both perceiver groups, there was no significant group differences overall.

Therefore, as with Experiment 4:1, these findings showed that people are capable of using other minds to infer external events that happened to them, only if others are not simulating a different mind. Even if people can discriminate different expressions, this is insufficient for perceivers to retrodict to the actual cause of the expression. In making their retrodictive judgements, retrodicting

perceivers performed similarly to perceivers who were merely classifying targets' expressions, offering little in the way of evidence to suggest that retrodicting and classifying perceivers engaged in different strategies in this experiment.

7.9 General Discussion

The results of Experiments 4.1 and 4.2 suggested that perceivers were better than chance in retrodicting from targets' behaviour to whether targets were looking at a positive or negative picture, from spontaneous and congruent reactions. In other words, when others are behaving naturally or simulating a behaviour congruent with what they are experiencing, people can use targets' minds as a lens onto external events that happened in the world. However, perceivers were systematically inaccurate at retrodicting from targets' incongruent behaviour to what they were actually looking at, suggesting that people are not capable of using other minds if others choose to simulate a different mind. This supports DePaulo and Morris (2004)'s proposition that people are credulous in interpreting signals from other people as being valid and truthful.

Although people are fooled by others' minds, people can discriminate between different types of behaviour. Specifically, they can discriminate between occasions in which others are behaving naturally and when others are simulating an expression. In relation to simulations, people can also discriminate between occasions where people are simulating a congruent or incongruent reaction. However, this ability to discriminate is not sufficient to make an accurate retrodictive inference of what really and truly caused the reaction. These findings are also consistent with DePaulo (1994)'s findings which suggest that even though

lies are judged less truthful than true statements, perceivers cannot see through the lies to know the true state. This could be attributed to perceivers relying on invalid behavioural cues, as suggested in classic lie detection studies (DePaulo et al., 2003; Bond and DePaulo, 2006; Vrij et al., 2008). When perceivers saw an incongruent reaction, they might have recognized it as incongruent. But, as an incongruent reaction may incorporate two different types of expression, perceivers may be confused as to which expression truly reflects targets' true inner state, thereby causing them to make inaccurate retrodictive judgements. Alternatively, targets may have displayed their cued expressions very convincingly to the extent that there are only minute differences between the actual and simulated inner state (Sporer and Schwandt, 2006, 2007), thus making it difficult for perceivers to identify the actual inner state.

The pattern of eye movements did not discriminate between congruent and incongruent simulations though they did discriminate between occasions when targets were behaving naturally and when they were acting upon a cue. Perceivers attended to the eyes and mouth equally for both spontaneous reactions to positive and negative pictures, presumably because there were no visible differences between spontaneous reactions to positive pictures and spontaneous reactions to negative pictures. As for congruent reactions, perceivers looked at the mouth longer for reactions to positive pictures than reactions to negative pictures, while the opposite was found for the eyes. This could be because there was more visible information at the target mouth region for congruent reactions to positive pictures but more visible information at the target eye region for congruent reactions to negative pictures. This explanation suggests that people attend to the region of the face which provides them with the most information, consistent with past studies

which found that the pattern of eye movements vary across different scenarios (Cassidy et al., 2014; Eisenbarth and Alpers, 2011; Pillai et al., 2012).

In making retrodictive judgements, the answer as to whether retrodicting perceivers use a different strategy than classifying perceivers is inconclusive. When perceivers were asked to infer whether targets were looking at a negative picture (Experiment 4:1), retrodicting perceivers (but not classifying perceivers) were sensitive to the negative and non-negative cause of targets' spontaneous reactions. But, when perceivers were asked to infer whether targets were looking at a positive picture (Experiment 4:2), both groups of perceivers were equally sensitive to the positive and non-positive cause of targets' spontaneous reactions. Spontaneous reactions are often very subtle and may be perceived as negative by default. As such, when classifying perceivers were asked to judge if targets' expressions were negative, perhaps they relied solely on the information available in targets' expressions. This could lead to them mis-classifying subtle spontaneous reactions to positive pictures as negative expressions, thus giving rise to a higher false alarm rate. But, when perceivers were asked to classify an expression as positive, classifying perceivers correctly rejected subtle spontaneous reactions to negative pictures as expressions which were not positive, thus, reducing the false alarm rate. These findings support Johnston et al. (2001)'s findings who found a mutual confusion between negative and neutral faces more frequent than confusion between happy and neutral faces. This is because a single feature of the face can be easily linked with happiness (Adolphs, 2002). When such a salient feature can be used to recognize the stimulus, there is no need for a complete visual analysis of the stimulus, unlike negative expressions (Fabre-Thorpe et al., 2001). As for retrodicting perceivers, perhaps they did not solely rely on the facial

behaviour of targets. Rather, upon viewing targets' expressions, perceivers might have retrodicted from other aspects of targets' behaviour to their inner states, before inferring what picture could have caused targets' reactions (Gallese and Goldman, 1998). Not only that, the pattern of eye movements across the duration of stimulus presentation differs slightly between both perceiver groups, suggesting that there might be a difference in perceivers' eye movements depending on what they were asked to do. These findings could be consistent with the possibility that retrodicting perceivers engaged in an elaborated and indirect form of processing while classifying perceivers engaged in a basic and direct form of processing, though the evidence to this effect is extremely tenuous.

Although these findings may suggest that perceivers in both groups engaged in different strategies, in most cases there were no significant group differences in any of the comparisons. With regards to an elaborated and indirect form of processing, retrodicting perceivers might have retrodicted from targets' behaviour to targets' inner states prior to inferring what picture they were looking at. But classifying perceivers might have also retrodicted from targets' behaviour to targets' inner states before judging the expression targets were displaying. Conversely, both perceiver groups might have engaged in a direct and simple form of processing in such a way that retrodicting perceivers directly inferred what picture targets were viewing from targets' behaviour, avoiding any inference of inner states, similar to classifying perceivers who selected the option that best represents targets' expressions without considering inner states.

In conclusion, people can use other minds as a lens onto external events that happened in the world, so long as targets are behaving naturally or simulating a

mind that is congruent with what they are experiencing. However, perceivers are not able to infer the true cause of targets' inner states, should those targets wish to simulate a mind which is different from what they were actually experiencing. Nevertheless, perceivers are sensitive to how others are behaving and whether they are behaving naturally or acting. As to whether retrodicting perceivers were using a different strategy than classifying perceivers, the findings remain inconclusive.

CHAPTER 8

Discussion

8.1 Research Overview

The face provides a rich source of information pertaining to one's identity (Burton et al., 1993; Blair et al., 2000) and emotions (Ekman, 1972; Izard, 1971; Tomkins, 1962). Studies investigating people's ability to recognise facial expressions of emotions have often asked people to match or label photographs of facial expressions according to their respective emotional terms. Although there are mixed findings as to whether the recognition of a particular set of basic emotions is universal cross-culturally (Biehl et al., 1997; Ekman et al., 1969; Ekman and Friesen, 1971; Jack et al., 2009; Matsumoto, 1992; Nelson and Russell, 2013; Wang et al., 2006), these studies have confirmed that people can reliably and consistently recognise emotions from facial expressions. Due to the limitations of these tasks, though, such as the artificial nature of the stimuli and the problem of an objective answer, the findings cannot be generalised to people's ability to recognise everyday subtle facial expressions. Since people can accurately recognise facial expressions of emotions, it has been proposed that the mind is signalled or embodied in one's behaviour, particularly facial expressions. If one's mind were embodied in their behaviour, do people still make inferences about inner states while inferring their emotions? As these tasks only require one to match, label or categorise photographs into piles associated with their respective emotional labels, perceivers can succeed by merely classifying their responses according to a

specific facial configuration (Adolphs, 2002), bypassing any form of inferences of inner states (Green, 2007; Gallagher, 2008; Smortchkova, 2017).

Since the recognition of facial expressions of emotions might not require any inferences of inner states, presumably because the mind is signalled in the face, are there any tasks which require one to make inferences of inner states based on targets' behaviour which are not facial expressions? One of the most prominent tasks which requires the inference of inner states is the False Belief task used in traditional "Theory of Mind" research. In these tasks, participants were asked to predict where a protagonist will look for an object (Wimmer and Perner, 1983; Wellman et al., 2001) or retrodict from the protagonist's behaviour to their false belief (Bartsch and Wellman, 1989; Robinson and Mitchell, 1995). One of the major limitations of this task is that the characters in these tasks are hypothetical entities. People do not usually infer whether hypothetical entities hold a false belief in everyday social situations; hence, these false belief tasks do not test people's ability to make practical inner state inferences in everyday social interactions. Instead, we usually make retrodictive inferences from another person's spontaneous behaviour. One such retrodictive inference would be the situational cause of a person's behaviour. We do not stop at inferring another's inner state, rather, we continue processing to identify the cause of the inner state. In an attempt to do so, people could verbally ask others about the root cause of their behaviour in order to understand them better. In the absence of being able to request a verbal explanation, is it nevertheless possible for people to infer the situational cause of the inner state and the behaviour?

The procedure for investigating retrodictive inferences has some useful features. Unlike past studies which used artificial stimuli (for example, false belief, emotion recognition), this method investigates people's ability to make inferences beyond one's inner states (such as inferences about an external event) in the social world. In addition, the procedure allows us to investigate if people make inferences about others' inner states when retrodicting from targets' behaviour to the situational cause of a behaviour. As the performance of retrodicting perceivers were compared against the performance of classifying perceivers (with the only difference in procedure being the content of the test question), any difference in performance between groups can be attributed to the task instructions and not to any other irrelevant factors. Even if the findings reported in this thesis did not clarify whether retrodicting perceivers inferred targets' inner states, perceivers' ability to infer the situational cause of target expressions indicates that people are able to make inferences beyond emotions from a target's dynamic, subtle and brief behaviour.

There are various causes precipitating a person's inner states, such as exposure to the emotions of a third party. Is it possible for people to accurately retrodict from a person's behaviour to the emotional expression of a third party? Chapter 4 investigated whether perceivers can make an accurate retrodictive inference from targets' behaviour to an emotional expression that they were viewing. If perceivers can accurately retrodict from target's expressions, it is assumed that perceivers were inferring targets' inner states before judging the situational cause of the behaviour (Teoh et al., 2017). However, perceivers were not explicitly asked to reason about targets' inner states and as such, an alternative view would be that perceivers might infer the situational antecedent of a

behaviour, without drawing any inferences of the target's inner states. If retrodicting perceivers were engaging in a simple and direct route of processing, then they would perform similarly to perceivers who were merely asked to classify target responses. Chapter 5 thus investigated not only whether perceivers were able to make accurate retrodictive inferences but also whether they were engaging in a simple or elaborated form of processing. Although people are exceptionally good at making various retrodictive inferences from targets' spontaneous behaviour (Cassidy et al., 2014; Pillai et al., 2012; Teoh et al., 2017), it is unknown as to whether people are accurate at making these inferences from simulated target expressions. Given that deception is a fundamental aspect of human social interaction (DePaulo et al., 2003), can people nonetheless accurately retrodict from others' simulated behaviour to infer what the target is actually experiencing?

Chapter 6, therefore, investigated whether perceivers can infer when targets were displaying a simulated incongruent expression, by testing their ability to infer whether targets were responding to a congruent or incongruent cue word-picture pairing. Even if perceivers were systematically able to judge that targets were displaying an incongruent expression, it is uncertain as to whether they can retrodict from the incongruent behaviour to the true, underlying situational cause of the target's behaviour. Hence, Chapter 7 investigated whether perceivers can accurately retrodict from targets' spontaneous, simulated congruent and simulated incongruent expressions to the type of picture that they were viewing. In other words, it investigated if perceivers can 'see through' targets acting in order to apprehend the actual state targets were experiencing. This chapter also investigated whether retrodicting perceivers engaged in an elaborated or simple form of processing in making their judgements.

In summary, the research asked the following questions: Were perceivers were accurate in retrodicting from targets' spontaneous behaviour to the emotional expressions that they were looking at? In making these retrodictive inferences, were perceivers engaging in a simple or elaborated strategy? Can perceivers guess if others were displaying behaviour that is incongruent with their true feelings? Were perceivers accurate at retrodicting from targets' simulated expressions to the type of pictures that they were actually looking at? In making these retrodictive inferences from targets' simulated behaviour, were perceivers engaging in a simple or an elaborated processing strategy?

8.2 SUMMARY OF RESULTS

8.2.1 Experiment 1- Can individuals guess what others were looking at, based on brief samples of their behaviour?

The aim was to investigate whether perceivers were accurate in making retrodictive inferences from targets' behaviour to the emotional expressions they were looking at. Targets were surreptitiously filmed and were either asked to memorise or imagine what the actor was feeling, while viewing photographs of emotional expressions. Perceivers were then asked to infer the emotional expression targets were looking at. Perceivers were significantly accurate at discriminating between reactions to all different emotional expressions when targets were asked to memorise the expressions. But, when targets were asked to imagine what the actor was feeling, perceivers were significantly accurate for all (except fear). Perceivers' accuracy differed depending on whether targets were asked to memorise the expression or imagine what the person in the picture was

feeling. Across both target conditions, perceivers were significantly better at judging reactions to happy, compared to surprised, anger, disgusted, fear and sad, regardless of whether targets were asked to memorise the picture or imagine what the target was feeling. They were also significantly more accurate at inferring reactions to sad than fear expressions. When targets were asked to memorise the expressions viewed, perceivers were significantly more accurate for sad than angry and disgusted. In terms of response times, perceivers responded faster for reactions to surprised expressions than happy, sad and angry expressions.

Interpretation

Perceivers were significantly above chance at inferring the emotional expressions that targets were viewing, despite the variability and subtlety of these expressions. Perceivers were accurate, presumably because targets spontaneously imitated the expressions that they viewed, regardless of whether they were asked to memorise the expression or imagine what the actor was feeling (Blairy et al., 1999; Lundqvist and Dimberg, 1995; Wood et al., 2016). Perceivers were also significantly better at making retrodictive inferences from happy expressions, as compared to other expressions. This could be because the spontaneous imitation of a happy expression involves only one distinctive feature such as a smiling mouth (Adolphs, 2002). In contrast, the imitation of negative emotions might be more difficult (Öhman et al., 2001), thereby resulting in a significantly less accurate performance among perceivers when observing targets who were looking at negative facial expressions.

Thus, perceivers were significantly accurate, although slightly better above chance, at retrodicting from targets' facial expressions to infer the emotional

expressions that those targets were looking at, implying that people are able to use the minds of others as a lens onto the emotional expressions of a third party who are not within their view. When perceivers viewed targets' expressions, they perhaps inferred the inner state of targets prior to inferring the emotional expressions they were looking at.

8.2.2 Experiment 2 – Differences in processing strategies between retrodiction and classifying based on facial expressions

Since Experiment 1 revealed that perceivers can use other minds as a basis to infer the emotional expressions viewed by targets, Experiment 2 aimed to replicate these findings. In addition, Experiment 2 also aimed to investigate whether retrodicting perceivers engaged in an elaborated form of processing which involved the inferences of inner states. Targets were surreptitiously filmed while they were viewing photographs of emotional expressions and had to choose the emoticon which best represents the expression viewed. Retrodicting perceivers were asked to judge the emoticon chosen by the targets while classifying perceivers were asked to select the emoticon which best represented targets' expressions. Retrodicting perceivers' performance, reaction times and eye movements were compared against that of classifying perceivers.

In Experiment 2.1, both retrodicting and classifying perceivers were slightly above chance in their responses, only when targets' expressions were most expressive. Group differences were manifested in two different ways, via the difference in the pattern of rank responses and reaction times. Retrodicting perceivers were most accurate for happy, followed by neutral, disgusted, angry, surprised and sad. Meanwhile, classifying perceivers were most accurate for

happy, followed by disgusted, neutral, surprised, angry and sad. In terms of reaction times, retrodicting perceivers took a longer time to respond than classifying perceivers.

In Experiment 2.2, both retrodicting and classifying perceivers were slightly above chance in their responses, but only when targets' expressions were most expressive. Differences were apparent in the pattern of rank reaction times between the two perceiver groups; otherwise there was an absence of evidence pertaining to differences between perceiver groups. Classifying perceivers responded fastest for happy, followed by sad, surprised, disgusted, neutral and angry. Retrodicting perceivers responded fastest for sad, followed by happy, neutral, surprised, angry and disgusted expressions.

Eye movement data revealed that perceivers looked at the eyes longer for least expressive target reactions to happy and surprised, as compared to sad and angry. But, they looked at the eyes longer for most expressive target reactions to angry and neutral expressions, as compared to happy, surprised, sad and disgusted. Gaze duration at the eyes was significantly longer for least expressive than most expressive target reactions to happy and surprised expressions. In relation to gaze duration at the mouth, perceivers looked longer at the mouth for most expressive target reactions to surprised than neutral and sad expressions. Moreover, gaze duration at the mouth was significantly longer for most expressive target reactions than least expressive target reactions, only for happy expressions. Finally, there was no significant group differences in the amount of dwell time at the target's eyes and mouth.

Interpretation

Overall, based on both experiments in Experiment 2, retrodicting perceivers were above chance level in inferring the emoticons chosen by targets. These findings imply that they were significantly accurate in their inferences of what emotional expressions targets viewed, hence, guiding their inferences of which emoticons targets chose. Classification perceivers, too, were significantly accurate in guessing which emoticons best represented target expressions. However, both groups of perceivers were accurate only when targets' reactions were expressive, suggesting that perceivers were only successful when the signal was relatively strong (Wood et al., 2016).

When retrodicting perceivers made their responses, were they engaging in a simple or elaborated processing strategy? Group differences manifested in different ways across the two studies. The inconsistency in group differences does not provide compelling evidence that there are categorical differences in the processing strategy used between groups. It would, therefore, be better to assume that both groups used similar processing strategy, whether simple or elaborated. Retrodicting perceivers might have engaged in the simple and direct route of processing similar to classifying perceivers, by choosing the emoticon which best represents targets' expressions without inferring targets' inner states. Conversely, both groups of perceivers might have engaged in an elaborated and indirect route of processing. Retrodicting perceivers might have inferred targets' inner states before inferring the emotional expressions that targets were looking at and, likewise, classifying perceivers might have inferred targets' inner states before selecting the emoticon which best represents targets' expressions.

Lastly, gaze behaviour at the eyes and mouth were not affected by whether perceivers were asked to retrodict or classify. Rather, the pattern of eye movements varied across the different levels of expressiveness and different types of expressions viewed (Eisenbarth and Alpers, 2011; Pillai et al., 2012).

8.2.3 Experiment 3 – Can individuals guess if others are simulating their expressions?

Experiment 1 and Experiment 2 demonstrated that people can use other minds to infer the emotional expressions that they were viewing. More broadly, these findings showed that people can use other minds as a lens onto external events that happened in the world. These expressions are spontaneous and genuine. In real-life, however, people are adept at deceiving others (DePaulo, 2003), where expressions are neither spontaneous or genuine. Therefore, this study aimed to investigate if people could infer whether targets were simulating an incongruent or congruent expression, by testing their ability to infer whether targets were reacting to an incongruent or congruent cue word-picture pairing. While targets were viewing a positive or negative picture, they were asked to display a congruent or incongruent expression as directed by a preceding cue word. Perceivers were then asked to judge if targets were responding to a congruent or incongruent cue word-picture pairing. The findings showed that perceivers were significantly above chance in guessing whether targets were reacting to an incongruent cue word-picture pairing, only when targets were viewing a negative picture; perceivers performed at chance level in judging whether targets were reacting to an incongruent cue word-picture pairing when

targets were viewing a positive picture. As for intensity ratings, perceivers rated incongruent reactions as significantly more expressive than congruent reactions, when targets were viewing negative pictures. Furthermore, perceivers rated congruent reactions to positive pictures significantly more expressive than congruent reactions to negative pictures.

Interpretation

Perceivers were, thus, able to identify if targets were simulating an incongruent expression only when targets were experiencing a negative image. This could be because incongruent expressions to negative pictures were stronger than congruent expressions to negative pictures, thus, providing a clearer signal to the perceiver than when targets were reacting to an incongruent cue word-picture pairing. When targets were looking at a negative picture, targets' simulation of a positive expression was not convincing, thereby leading perceivers to infer that targets were displaying an incongruent expression. But, when targets were reacting to a positive picture, their simulation of a negative expression was very convincing to the point that perceivers were fooled by their expressions, resulting in their inability to judge whether targets were responding to an incongruent or congruent cue word-picture pairing. These results are consistent with Porter and ten Brinke's (2008) study which suggested that negative feelings were more difficult to falsify. Overall, people were significantly accurate at judging if others were simulating an incongruent expression, only when they were experiencing something negative. This could be done through a process of retrodiction whereby perceivers first infer the true inner state of targets, before judging whether they were reacting to an incongruent or congruent cue word-picture pairing.

8.2.4 Experiment 4 – Can individuals retrodict from simulated expressions?

Given that people are capable of inferring whether others were simulating an incongruent expression when they are experiencing a negative situation, Experiment 4 examined whether perceivers were able to make accurate retrodictive inferences from these simulated expressions. As with Experiment 2, this experiment also examines whether retrodicting perceivers inferred targets' inner states through an elaborated form of processing. Perceivers were shown targets' spontaneous, simulated congruent and incongruent reactions to negative and positive pictures. Retrodicting perceivers were asked to judge if targets were looking at a negative picture (Experiment 4.1) or a positive picture (Experiment 4.2). Classifying perceivers, in contrast, were asked to judge if the target expression was negative (Experiment 4.1) or positive (Experiment 4.2).

In Experiment 4.1, retrodicting perceivers were significantly above chance in judging whether targets were looking at a negative picture for spontaneous and congruent expressions. Classifying perceivers were also significantly above chance in inferring if targets' expressions were negative, but only for congruent expressions. Both perceiver groups were significantly below chance in their responses for incongruent target reactions. Although both perceiver groups were inaccurate for incongruent target reactions, their pattern of responses discriminated between the different types of reactions. Firstly, perceivers' responses discriminated between simulated congruent and incongruent expressions. When targets simulated a negative expression, perceivers judged that targets' expressions were negative significantly more frequent when they were actually looking at a negative picture (congruent expressions) as compared to a

positive picture (incongruent expressions). Next, perceivers' responses also discriminated between spontaneous and congruent expressions. When targets were viewing a negative picture, perceivers stated that targets' reactions were negative significantly more frequent when they were simulating a congruent expression as compared to a spontaneous expression. Conversely, perceivers stated that targets' expressions were negative more frequent for spontaneous reactions than congruent reactions to positive pictures.

In terms of reaction time, perceivers took significantly longer to respond to spontaneous than congruent and incongruent reactions. As for expressiveness ratings, target reactions to negative pictures were judged to be more expressive than target reactions to positive pictures, for both spontaneous and incongruent reactions. Meanwhile, target reactions to positive pictures were judged to be more expressive as compared to negative pictures, for congruent reactions. When targets were looking at a negative picture, incongruent reactions were judged to be the most expressive, followed by congruent and spontaneous expressions. But, when targets were looking at a positive picture, congruent reactions were rated the most expressive, followed by incongruent and spontaneous expressions. As for group differences, perceivers in the retrodicting group performed significantly better than perceivers in the classifying group. But, there were no significant group differences for reaction times.

In Experiment 4.2, both retrodicting and classifying perceivers were significantly above chance in their responses for spontaneous and congruent reactions. In contrast, both groups were significantly below chance in their responses for incongruent reactions. As with Experiment 4.1, perceivers'

responses discriminated between different types of target reactions. First of all, perceivers' responses discriminated between congruent and incongruent reactions. When targets simulated a positive expression, perceivers stated that targets' expressions were positive more frequently when they were truly viewing a positive picture (congruent expression), as compared to when they were viewing a negative picture (incongruent expression). When targets simulated a negative expression, perceivers stated that targets' expressions were positive significantly more often when targets were viewing a positive picture (incongruent expression), as compared to when they were viewing a negative picture (congruent expression). In addition, perceivers' responses also discriminated between spontaneous and congruent reactions. When targets looked at a positive picture, perceivers judged that targets' expressions were positive more often for congruent reactions than spontaneous reactions to positive pictures.

As for reaction times, perceivers took significantly longer to respond to congruent target reactions to negative pictures than congruent target reactions to positive pictures. In addition, perceivers also took a longer time to respond to spontaneous reactions than congruent and incongruent reactions.

In terms of expressiveness ratings, perceivers rated incongruent reactions as the most expressive, followed by congruent and spontaneous reactions. As with Experiment 4.1, perceivers rated congruent target reactions to positive pictures as more expressive than congruent target reactions to negative pictures. Meanwhile, perceivers rated incongruent target reactions to negative pictures as more expressive than incongruent target reactions to positive pictures. For target reactions to negative pictures, perceivers rated incongruent expressions more

expressive than congruent and spontaneous expressions. In contrast, when targets were looking at a positive picture, perceivers rated congruent reactions as the most expressive, followed by incongruent and spontaneous reactions.

In relation to the pattern of eye movements, perceivers looked longer at the eye region for congruent reactions to a negative picture than congruent reactions to a positive picture. Perceivers also looked at the eyes significantly longer for spontaneous than congruent reactions to positive pictures, but not for reactions to negative pictures. As for gaze duration at the mouth region, perceivers looked significantly longer at the target's mouth for reactions to positive than negative pictures, only for congruent reactions. Perceivers also looked longer at the target's mouth for spontaneous reactions than congruent reactions to negative pictures. They also looked longer at the target's mouth for congruent than spontaneous reactions to positive pictures. There were, however, no significant group perceiver differences in terms of accuracy, reaction times or eye movements.

Interpretation

The findings from both studies indicate that retrodicting perceivers were above chance level in making retrodictive judgements from spontaneous and congruent expressions. In contrast, classifying perceivers' responses for spontaneous target reactions were at chance level when judging whether targets' reactions were negative but were above chance when judging whether targets' reactions were positive. Classifying perceivers also performed above chance level at judging whether targets' reactions were negative or positive from congruent reactions. Both groups of perceivers performed significantly below chance for incongruent reactions, suggesting that they were fooled by targets' simulations.

However, the pattern of responses, reactions times, expressiveness ratings and eye movements, to a certain extent, discriminated between different types of reactions. But, this discrimination was not sufficient to make an accurate retrodictive inference of what truly caused the target's expression. These findings are consistent with DePaulo (1994)'s findings, suggesting that even though perceivers rated lies as less truthful than truths, they were not accurate at explicitly discriminating between lies and truths. Thus, perceivers are only able to utilise other minds to infer what really caused an expression if targets were behaving spontaneously or simulating a behaviour that is congruent with what they were experiencing. However, it seems people are unable to use other minds to infer the precipitating external event, if others choose to simulate a behaviour which is different from what they were truly experiencing.

When perceivers were asked to retrodict, their responses, reaction times and eye gaze did not differ from classifying perceivers, implying that both groups engaged in the same processing strategy. As with Experiment 2, perceivers in both groups might have engaged in a simple or elaborated strategy. In relation to the former, retrodicting perceivers might have responded by selecting the option that best represented targets' expressions, similar to classifying perceivers. As for the latter, retrodicting perceivers might have inferred targets' true inner states prior to inferring what targets were looking at; and classifying perceivers might have inferred targets' inner states prior to judging whether targets' expressions were negative or positive.

In summary, people are able to use other minds as a lens onto external events that happened in the world, only if targets' behaviour was spontaneous or

congruent with what they were truly experiencing. But, there is insufficient evidence to conclude that retrodiction involves the inference of inner states in this case.

8.3 PRACTICAL AND THEORETICAL IMPLICATIONS

8.3.1 Retrodictive Inferences

People are above chance level in their ability to retrodict from another person's behaviour to infer the event that precipitated the behaviour; they are able to make these inferences when the person they are observing is either behaving spontaneously or simulating a behaviour which is congruent with the event they were experiencing. These findings are consistent with previous retrodicting studies (Cassidy et al., 2014; Pillai et al., 2012; Teoh et al., 2017) which demonstrated that people are systematically accurate at retrodicting from a person's behaviour to the situational cause. These findings point to the conclusion that individuals are able to access an external event in the world through the lens of another person's mind (as signalled in their behaviour). However, people are not accurate when observing another person simulate a behaviour which is incongruent with the event they were experiencing. As there were no reliable findings suggesting any differences, depending on whether perceivers were asked to classify the target's expression or infer the situational cause of the target's expression, there is no evidence to suggest the different kinds of test question provoked different kinds of processing in the perceivers.

When perceivers retrodict, it is assumed that they engaged in an elaborated and indirect form of processing (as shown in Figure 8.1 below), where they first infer the target's inner state (proximal cause) before inferring the antecedent event that caused the inner state (distal cause) that in turn caused the consequent behaviour (Teoh et al., 2017).

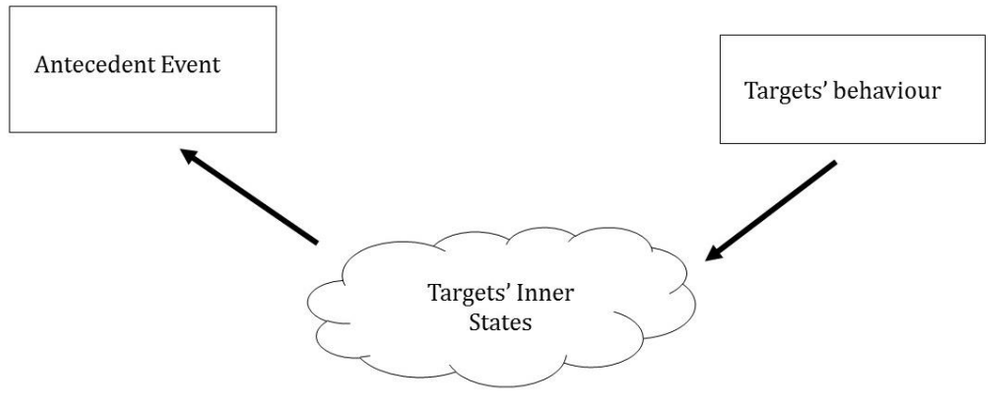


Figure 8.1. This elaborated and indirect form of processing implies there is a sequential mental operation involved in making retrodictive inferences, such that perceivers first infer targets' inner states before inferring the antecedent event.

If people were to engage in a serial form of processing, what is the mechanism behind this process? How do people make inferences about targets' inner states from their behaviour? Having inferred targets' inner states, how do people then infer the event that caused the behaviours? Perceivers' inferences of inner states such as emotions from facial expressions can be interpreted as a basic form of retrodiction (Gallese and Goldman, 1998) where perceivers retrodict from targets' behaviour to their inner states through a process of simulation. In a simple retrodiction task which requires perceivers to recognise targets' inner states,

perceivers first recognise the inner state displayed by targets by creating a similar expression in themselves. There are many ways in which perceivers can recognise inner states from faces via simulation (Goldman and Sripada, 2005) such as generate-and-test simulation, reverse simulation, reverse simulation with "as-if" loop and unmediated resonance.

The generate-and-test model (as seen in Figure 8. 2) assumes that upon seeing a target's facial expression, the perceiver starts hypothesizing an emotional state that could be the possible cause of the target's expression. The perceiver then generates the hypothesized emotion in his or her own mind which might include reproducing this facial expression. Then, the perceiver attempts to match the hypothesized emotion, through a learned association between proprioceptive and visual representations of the same facial configuration. If the resulting facial expression matches the observed expression, then the hypothesized emotion is confirmed and the perceiver attributes the hypothesized emotion to the target. If the resulting facial expression does not match the observed expression, then the perceiver would hypothesize a different emotion.

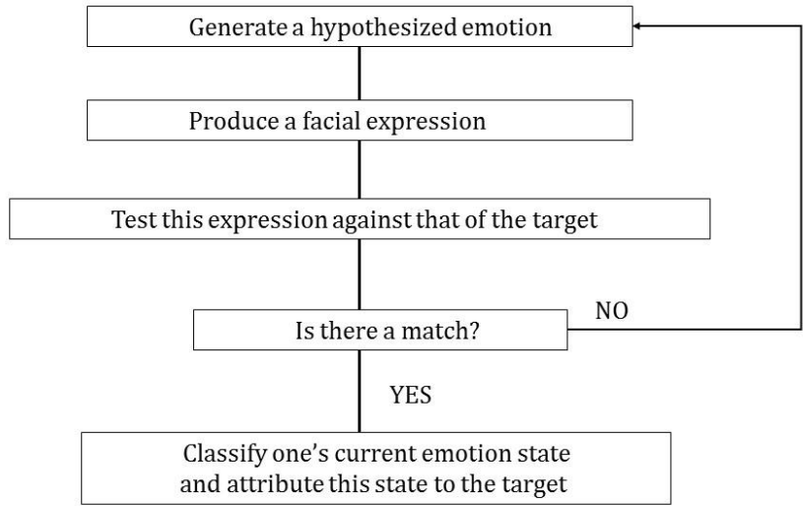


Figure 8.2. Generate-and-test simulationist model of emotion recognition.

In relation to the example of observing a smile, the perceiver first hypothesizes the emotion that could be the cause of the facial expression such as *happy* (inner state). The perceiver then generates a happy emotion in his or her own mind which might even include reproducing a smile. If the perceivers' smile or instruction to produce this smile matches the observed smile, then the hypothesized emotion (in this case, happy) is confirmed and consequently, the perceiver attributes the happy emotion to the target.

As for the reverse simulation model (Figure 8.3), it supposes that the perceiver first imitates targets' expressions in a covert manner. As changes in facial musculature produce traces of the corresponding emotion, these traces are then classified into their respective emotion category. The perceiver then attributes the emotional state to the target.

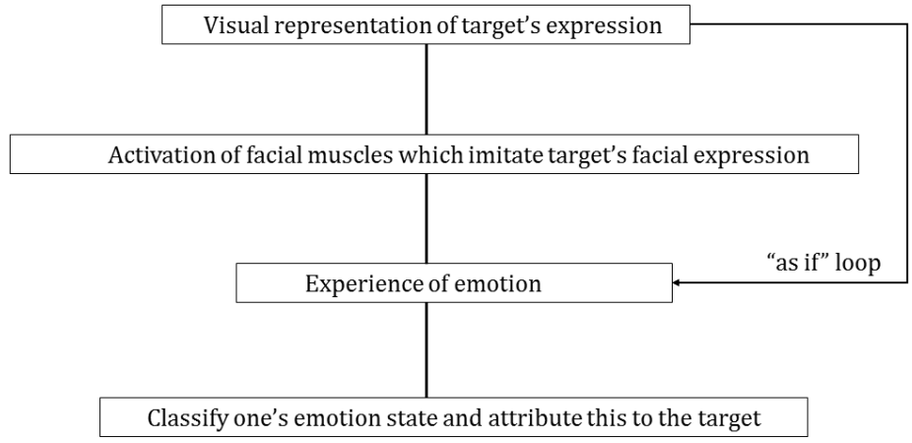


Figure 8.3. Reverse simulation model and the modification of the reverse simulation model with an "as-if" loop.

For example, upon seeing a smile, the perceiver would imitate the smile in a covert manner. As these changes in the facial musculature of targets' imitation produce traces of the corresponding emotion (such as *happy*), the perceiver then classifies targets' smile as signalling a happy emotion. Finally, the perceiver attributes the happy emotion to the target.

In a modified version of this model, the reverse simulation model with an "as-if" loop states that the production of facial expressions via changes in facial musculature can be bypassed. Rather, viewing facial expressions alone could simulate how perceivers would feel if they had made the facial expression displayed by the target. This link between the visual representation of facial expressions and a somatosensory representation of what it would feel like to make that expression would lead to the activation of the emotion, corresponding to the

targets' expression. This emotion is then classified and attributed to the target. For instance, upon viewing a target's smile, the perceiver simulates how he or she would feel if he or she were smiling. Due to a direct pathway between the visual representation of the target's smile and the perceiver's somatosensory representation of the smile, the corresponding emotional category (in this case, happy) is activated. This happy emotional state would consequently be attributed to the target.

Lastly, the unmediated resonance model states that the perception of the target's expression triggers the activation of the same neural substrate of the corresponding emotion. This model is based on the assumption that the production of an emotion in the perceiver is transmitted to a cognitive centre which recognizes the emotion, leading to an overt classification as the corresponding emotion. For example, viewing a smile would trigger the recognition of a happy emotion in perceivers. The production of this happy emotion would then lead to an overt categorization of the smile as a happy emotion.

Since perceivers come to the understanding of targets' emotions by internally replicating their simulations without any explicit, reflective mediation, perceivers were not required to reason about others and use this reasoning to understand others' minds. While perceivers might be able to recognise targets' inner states by means of simulation, this process in itself, is insufficient to infer the situation which precedes the inner state. Instead, perceivers would need to infer targets' inner states and engage a second step which requires them to draw upon their knowledge of the possible situations that can give rise to the inner state in

question, to arrive at a judgement of the event that caused the behaviour. For instance, upon observing a smile, the perceiver might reproduce the smile that feels like a happy emotion, and then draw upon his or her knowledge about the possible situations that can give rise to a happy emotion (for example, looking at a positive picture). Subsequently, the perceiver infers that the target is smiling because he or she was viewing a positive picture.

While the mechanisms involved in this elaborated form of processing might explain the performance of retrodicting perceivers, how would this form of processing explain the performance of perceivers who classified target expressions? As with retrodicting perceivers, classifying perceivers might recognise and infer the inner states of targets through a process of simulation. After inferring targets' inner states, classifying perceivers then select the option that best represents targets' inner states.

According to this account, perceivers inferred the inner states of targets and when perceivers were asked to identify the cause of the target's facial expression, we therefore suppose that perceivers inferred the target's inner state as a preliminary step to inferring the situational event experienced by the target that caused them to have the inner state. Although it might seem improbable that perceivers made a direct inference from targets' behaviour to the situational event, this alternative must be considered since perceivers were not explicitly asked about targets' inner states: According to the simple and direct route of processing, perceivers retrodict directly from targets' behaviour to an antecedent event, bypassing any inference of inner states (as shown in Figure 8.4 below).

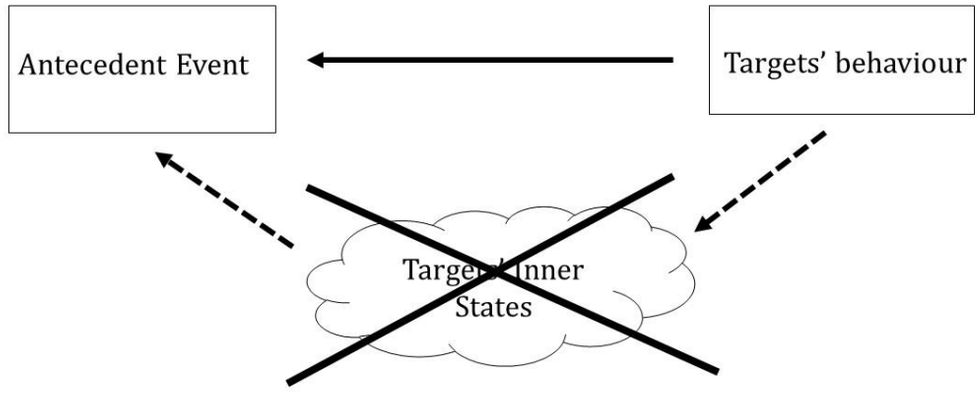


Figure 8.4. Simple and direct form of processing where perceivers retrodict from targets' behaviour to the event without inferring the targets' inner states.

Given that our minds are signalled in our facial expressions (Gallagher et al., 2008), one may argue that external events are also signalled in our expression to the extent that people would be able to match a specific facial configuration with a corresponding event that happened in the world according to a system of behavioural rules (Povinelli and Vonk, 2003; Perner and Ruffman, 2005), without inferring the inner states of targets.

If perceivers made inferences directly from observed target behaviour to the antecedent situational cause of that behaviour (without inferring target inner states), what would be the mechanism underlying this form of processing? If it is possible for perceivers to automatically recreate the target's inner state in their own mind (Goldman and Sripada, 2005), it might also be possible for perceivers to automatically recreate a given situation in their own mind. For example, upon viewing a smiling mouth, retrodicting perceivers might recreate the smile and thereby identify the corresponding situation in this case. They would effectively

infer that targets were smiling because targets were looking at a positive picture, without needing to infer that the target is feeling happy. As for classifying perceivers, they might simulate the expression viewed, for example, a smile, and choose the option that best represented what they were simulating, without having to infer that the target is feeling happy.

If perceivers were merely matching targets' behaviour to the situation, how does this relate to simulation? Presumably, if it is possible for perceivers to automatically recreate an emotion in the perceivers' mind, it might also be possible for perceivers to automatically recreate a given situation in the perceivers' mind using a simple form of retrodiction. It might very well be the case that the simulationist models of recognising emotions from facial expressions (Goldman and Sripada, 2005) could be extended to recognising situational antecedents from facial expressions in a similar manner. The process in which perceivers can arrive at judgements of situational antecedents by simulation, without inferring targets' inner states, is described below:

In the generate-and-test model, perceivers could generate a hypothesized situation instead on a hypothesized emotion, upon viewing targets' expressions. Perceivers then proceed to enact that situation which might include the production of the expression that corresponds with the situation. If the resulting facial expression matches the expression in the target, then the hypothesized situation is confirmed and the perceiver attributes the situation to the event that targets were reacting to (Figure 8.5).

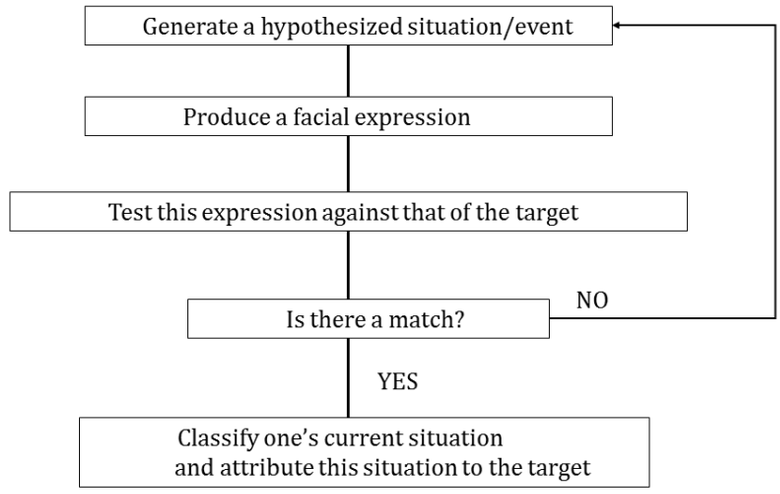


Figure 8.5. Generate-and-test model of a simple retrodiction model in retrodicting from targets' behaviour to the antecedent event, without inferring targets' inner states.

For instance, when a perceiver observes a smile, the perceiver hypothesizes that it is a reaction to a positive picture and proceeds to enact an expression which corresponds to a reaction of a positive picture. If the reproduction of this expression matches the observed expression, then the hypothesized situation (in this case, a positive picture) is confirmed and the perceiver infers that the expression is a reaction to a positive picture.

As for the reverse simulation model, the perceiver mimics the observed expression which could produce traces of the situation which could have led to the expression. The perceiver would then classify the corresponding situation and consequently, classify the observed expression as being expressive of the same situation (Figure 8.6).

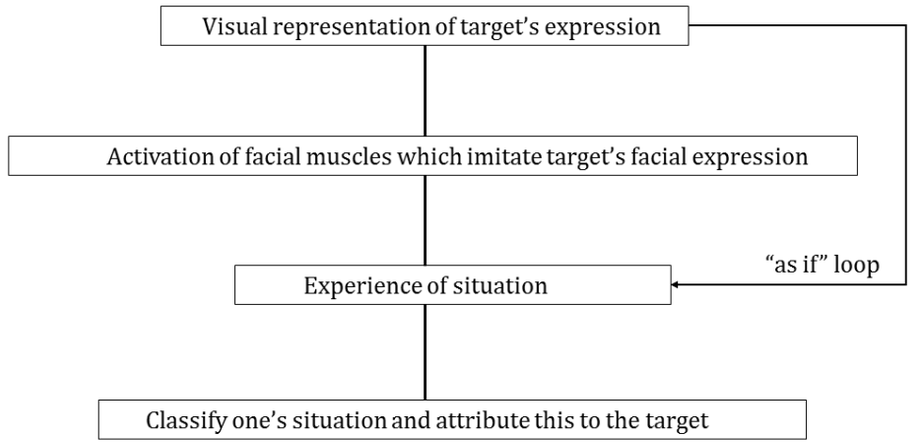


Figure 8.6. The reverse simulation model of retrodictive inferences.

For example, on viewing a smile, the perceiver mimics the smile which produces traces of a corresponding situation such as a positive picture. The perceiver would then classify the expression as a form of smile expressed when one views a positive picture. The perceiver then classifies the observed face as a reaction to a positive picture.

According to a modified version of the reverse simulation model (reverse simulation with "as-if" loop), perceivers would proceed directly to the somatosensory representation of what situation could have caused the facial expression and hence, bring about the corresponding situation. This modified version of the reverse simulation model states that there is no need to imitate the observed expression. For example, upon seeing a smile, perceivers would elicit a somatosensory representation of a certain situation that could have caused the smile and bring about the corresponding situation (in this case, a positive picture). The perceiver, then infers that the expression was a reaction to a positive picture.

The fourth model, unmediated resonance model, supposes that the observation of targets' expressions would trigger the activation of the same neural substrate associated with the situation. It is based on the assumption that the production of this situation is transmitted to a cognitive centre which recognizes the situation as the antecedent event of the expression, leading to a classification of the expression as the corresponding situation. As such, the observation of a smile would trigger the activation of the same neural substrate associated with looking at a positive picture.

Thus, the simulationist models of face-based emotion recognition can be extended to retrodicting from facial expressions to the antecedent event by supposing that perceivers automatically re-create the situation in self, without re-creating the inner state. However, it is implausible to suppose that retrodicting perceivers recreate a specific situation associated with a given facial expression. A particular expression can be caused by many different situations; conversely, a particular situation can give rise to many different expressions. In other words, expressions do not map simply and reliably onto situations. Therefore, it seems improbable that the perceiver would be able to directly infer the causal situation on observing target behaviour in the absence of additional information (or in the absence of making additional inferences). One such additional inference that could guide the inferences of the situational cause of the target's behaviour is an inference of the target's inner state.

Supposing that a perceiver observes a target smiling. There are many situations which would give rise to a smile. For instance, one might be smiling

because they received a compliment, received a gift, obtained good grades, solved a difficult problem or as a signal that they are looking at another person who is smiling. These situations, in turn, cause different corresponding inner states such as happy, delighted, pride, relief or flirtatious. If a perceiver directly perceives the situational cause of the target's behaviour, they would have to choose their answer from a wide range of possible situational causes. It would be very unlikely that they would obtain the right answer in the first instance; rather, it would be highly likely that their guessing would lead to the wrong answer. For instance, they might mistakenly guess that the target was smiling because he/she obtained good grades when in actual fact, the target was smiling because he/she managed to solve a difficult problem. But if perceivers inferred the target's inner states accurately, this would at least narrow down their range of choices of situational causes to those which are only related to the inner state, which in turn will increase their likelihood of giving a correct answer. For example, if perceivers inferred that targets were smiling because the target felt relieved, they would most probably infer that targets were smiling because they managed to solve a difficult problem. Hence, the inference of inner states would guide perceivers to make an accurate inference of the preceding situation.

8.3.2 Spontaneous and simulated target expressions

When targets simulate their expression or behaviour, are their true minds still transparent to perceivers? In order to retrodict effectively from targets' behaviour, perceivers must have the ability to see through others' simulations to infer whether they were simulating an expression that is congruent or incongruent with their experience. Are people's simulations convincing to the point that others are unable to infer whether they are simulating an incongruent expression? The

findings suggest people can only discriminate between congruent and incongruent expressions when the target is experiencing something negative. When the target is experiencing a negative situation, but simulating a positive expression, their simulations are not convincing, which consequently provides clues to others that they were simulating an expression. Conversely, when one is simulating a negative expression while experiencing something positive, their simulated expression is very convincing, potentially resulting in their true inner states being concealed from others. In short, people's behaviour reveals whether they were simulating a congruent or incongruent mind, only when they are experiencing something negative.

Although targets' minds are readable to others in the sense that others can infer if they were simulating an incongruent expression, are their minds accessible to the extent that others can infer the true antecedent event of their behaviour? Perceivers were accurate at inferring the antecedent event when others were displaying a behaviour that is congruent with what they were experiencing, implying that people's minds are transparent to others when they are simulating a behaviour that is congruent with the situation experienced by them. Conversely, people are systematically inaccurate and therefore fooled by targets' simulations when they were simulating an incongruent expression, implying that people's minds are not accessible to others when they are simulating an incongruent expression. Despite these findings, perceivers' responses discriminated between the different types of targets' expressions, implying that people's simulations (but not inner states) are detectable by others to some degree.

In summary, when people are behaving spontaneously or simulating an expression that is congruent with what they were experiencing, their minds are signalled in their behaviour and are interpretable to others, allowing a window onto events that happened in the world. However, when people choose to fool others by simulating an incongruent expression, they are largely successful. Although their incongruent behaviour might be detected by others, this is insufficient for others to infer the antecedent event, presumably because people's incongruent behaviours do not provide enough information about the true inner states of targets. While past studies have investigated whether people are adept at fooling others by means of verbal lies (DePaulo, 1994) or emotional deception (Ekman et al., 1988; Porter and ten Brinke, 2008), the findings in this thesis suggest people's simulations of an incongruent expression are sufficiently effective to conceal their true inner state and the actual event that they experienced.

8.3.3 Gaze behaviour

Facial expressions provide a rich source of information about the person's inner states. A smiling mouth may indicate that one is feeling happy while a frown may indicate that one is feeling annoyed. If people can make accurate inferences from facial expressions, is it achieved by directing attention towards a particular facial feature or to multiple facial features? Do perceivers' pattern of eye movements differ depending on the type of expression they are viewing? On viewing posed photographs of facial expressions, people seem to gaze longer at the mouth region compared to the eyes (Kirchner et al., 2011). Baron-Cohen et al. (1997), on the other hand, suggest that looking at the whole face is important for inferring basic emotions but that looking at the eyes in particular is most important when inferring complex emotions. Using a Bubbles technique that randomly

samples visual information contained in a stimulus such that a different subset of information is available at each trial, Blais et al. (2012) attempted to evaluate the most informative facial cue involved in discriminating between the facial expressions of the six basic emotions, pain and neutral. They found that the mouth was the most important facial cue for both static and dynamic stimuli (Blais et al., 2012). Similarly, Pillai et al.'s (2012) findings demonstrated that perceivers tended to look longer at the mouth than the eyes of targets when making retrodictive inferences. These findings suggested that perceivers attended to a specific facial feature (the mouth) across various different facial expressions.

However, the findings from this thesis showed a rather different pattern, which is that perceivers' gaze pattern differed depending on which expression the target was viewing and how intensely they spontaneously imitated the expression. While one might think that participants will attend to the mouth longer for happy expressions (Hanawalt, 1994), the pattern of eye movements differed depending on the expressiveness level of the reaction: Perceivers gazed longer at the eyes for happy and surprised target expressions, as compared to sad and angry, when they were viewing the least expressive target reactions; when perceivers viewed the most expressive target reactions, they gazed at the eyes longer for angry and neutral, as compared to the rest of the expressions. Perceivers also looked longer at the target's eyes for the least expressive reactions than most expressive reactions to happy expressions. As for gaze duration at the mouth region, perceivers looked longer at the mouth for the most expressive target reactions to surprised than neutral and sad expressions. Furthermore, perceivers gazed at the target's mouth significantly longer for most expressive reactions than least expressive reactions to happy expressions. These findings extend previous studies

(Eisenbarth and Alpers, 2011; Schurgin et al., 2014) in suggesting that the pattern of eye movements was not only affected by the type of reaction viewed, but was also affected by the intensity of the target's expression.

Furthermore, the pattern of perceivers' eye movements also discriminated between spontaneous and simulated target expressions. Perceivers looked at the eyes significantly longer for spontaneous target reactions to positive pictures than congruent target reactions to positive pictures. In addition, perceivers' gaze duration at the target's mouth was significantly longer for congruent target reactions to positive pictures than negative pictures, presumably because the target's mouth was more informative for positive expressions. Moreover, perceivers also looked longer at the target's mouth for congruent reactions than spontaneous reactions to positive pictures. When the target's mouth was expressive, perceivers tended to attend to the mouth. However, when the target's mouth was not expressive, perceivers tended to gaze longer at the target's eyes. As such, people's gaze direction is affected by which part of the face is most expressive (Eisenbarth and Alpers, 2011; Schurgins et al., 2014). Since these findings showed that there are differences in perceivers' eye movements depending on whether or not they are observing spontaneous and simulated target expressions, it is apparent that the way perceivers viewed posed expressions is different from the way they view spontaneous expressions. As such, if posed expressions are used as stimuli in experiments, any findings pertaining to the visual strategy used in social situations might not be generalisable to real-life contexts in which it is necessary for people to interpret spontaneous expressions.

Although Calvo et al. (2013) discovered that the mouth and the eyes of fake smiles received more fixations and longer dwell times than genuine smiles, these findings were not replicated in this thesis where there were no significant differences in the pattern of eye movements between simulated congruent and incongruent expressions. However, this could be due to the fact that Calvo et al.'s study utilised static stimuli while dynamic stimuli were used in this study. For static stimuli, the apex of incongruent expressions might be more expressive than the apex of the congruent expressions, leading to longer dwell times and more fixations for incongruent than congruent expressions. On the other hand, dynamic stimuli have a temporal characteristic in which a different amount of visual information is available at different points in time for both types of expressions. Although a different amount of visual information is available at different time points for both expressions, overall, they might have the same amount of information available. As a result, perceivers' gaze behaviour was similar for these expressions. Once again, the discrepancy in gaze behaviour depending on the type of stimuli being viewed highlights the importance of using ecologically valid stimuli that resembles real-life phenomena.

Finally, the findings in this thesis did not reveal any significant differences in perceiver gaze behaviour depending on whether they were instructed to infer which expression targets were looking at or instructed merely to classify the target's expression. These findings thus offer no evidence to suggest retrodicting and classifying perceivers employed different processing strategies. To sum up, perceivers' pattern of eye movements varied across different types and intensities of target expressions. These findings suggest that future studies investigating perceiver eye movements should use target stimuli which are spontaneous and

ecologically valid, in order to produce findings which are generalisable to gaze behaviour in everyday social interactions.

8.4 LIMITATIONS

Despite the advantages of this paradigm in presenting ecologically valid stimuli and fulfilling the "truth condition" (West and Kenny, 2011), there are also some limitations. Firstly, the studies in this thesis utilised a forced-choice design, similar to emotion recognition studies (Ekman et al., 1969, 1972; Baron-Cohen et al., 1996, 1997). The use of a forced-choice design in emotion recognition studies has been widely criticised (Nelson and Russell, 2013; Russell, 2003), for it constrains the way that perceivers respond and might give the impression that perceivers are a good deal more accurate in interpreting facial expressions than would be the case in real life. This follows in that perceivers' responses may not reflect their exact answers, but a close approximation to their exact answer. In real life, people are not asked to choose from a range of responses, rather, they have to respond freely. As such, if people are asked to make inferences about what happened to others using a free response method, they might not be able to produce the exact, correct answer. Hence, the use of forced-choice response might overestimate perceivers' ability to make these inferences in daily life. Nevertheless, their accurate performance in these studies indicates that even if their spontaneous answers would have differed from the choices that they selected, their recognition of the correct option in the context of incorrect alternatives suggests that perceivers are, to a certain extent, sensitive to what happened to others.

A further limitation concerns gender imbalance in the sampling. A majority of perceivers recruited for these tasks were female (Experiment 3, Experiment 4.1, Experiment 4.2). It has been suggested that females outperform males at identifying facial affect (Campbell et al., 2002; Collignon et al., 2011; Thayer and Johnson, 2000) but this also depends on the type of emotion viewed. Women are found to be better at recognising fear and sadness (Mandai and Palchondhury, 1988) while men are superior at recognising anger (Rotter and Rotter, 1988). In a recent meta-analysis by Thompson and Voyer (2014), there was a significant but small gender difference where females tended to outperform males on tasks assessing the ability to recognise non-verbal emotional displays, moderated by the type of emotion, age and gender of actors. However, past studies on retrodiction (Pillai et al., 2012; Teoh et al., 2017) suggest that people are systematically accurate at inferring the antecedent event of a situation, even when gender is matched.

Another limitation of this research is that a majority of targets and perceivers (except in Experiment 1) were British. Some studies have shown that people of different cultures vary in how they recognise emotional expressions (Matsumoto et al., 2008; Jack et al., 2009, 2012). Participants from Germany rated angry expressions as angrier, compared to Greek participants. In contrast, Greek participants rated sad expressions as sadder than German participants (Hareli et al., 2015; Hess et al., 2016). Moreover, some studies indicate that cultural differences exist with respect to how perceivers interpret a targets' face depending on whether it is viewed in isolation or in the context of others (Masuda et al., 2008). Japanese participants perceived facial expressions of emotions as more intense when targets were surrounded by people who showed the same emotion as the targets but this was not the case for Western participants, suggesting that Japanese

(and not Western) participants attended to the group's emotions and were affected by them (Masuda et al., 2008).

While the findings in this thesis have demonstrated that people in South East Asian cultures (Experiment 1) and Western participants (Experiment 2) are able to infer the antecedent event from spontaneous target expressions, it is uncertain as to whether cultural differences exist in people's ability to retrodict from simulated behaviour, and whether they employ different strategies. Thus, a heterogenous group of participants, consisting of various cultures should be included in order to draw universal conclusions about the process of retrodiction.

Although the stimuli presented capture the types of expressions we encounter in daily life, the experiment was nevertheless conducted in a controlled lab environment. Perceivers viewed targets expressions on the laptop screen and made their judgements with the knowledge that it is an experiment. This setting within a laboratory environment does not capture the features that exist in everyday social scenarios where the perceiver and target communicate face-to-face. Moreover, in Experiment 3 and Experiment 4, simulated congruent and incongruent expressions were obtained by asking participants to pose a particular expression. As these expressions are posed, targets' simulations might not represent how they would simulate an expression under real-life circumstances. In real-life, people would have a strong motivation for simulating their expressions and might therefore tend to exaggerate their simulations, providing clues to perceivers to try to see through the expressions. Alternatively, targets might be motivated to simulate a more natural expression, further concealing their true inner state to targets, making it even more difficult for perceivers to see through

the simulated expression. Likewise, perceivers did not have a strong incentive to see through targets' simulation whereas in real-life, people might be more highly motivated to infer if others were simulating, especially in high-stake circumstances such as negotiations or interrogations.

Lastly, the use of photographs from the IAPS database might be unfit for eliciting the intended emotions in targets. Just because the photographs are rated in terms of their valence where a low valence represents a negative picture and a high valence represents a positive picture, it does not necessarily follow that the pictures would elicit a corresponding emotion in targets. The positive pictures used might not have elicited as strong a positive emotion in targets, in comparison with the strength of the negative emotion elicited by the negative pictures. This is because the positive pictures used were mainly pictures of pleasant scenery (e.g., beach, ocean, fireworks) or animals (e.g., rabbits, kittens, baby seals) which targets might perceive as pleasant but not necessarily feel happy while looking at them. On the other hand, negative pictures (e.g., dirty toilet, vomit, car accident, riot) might have been perceived as unpleasant and at the same time, elicit a negative emotion. If so, it would have been easier for targets to simulate an incongruent expression when they were viewing a positive picture, thus, more easily preventing targets from seeing through their simulations compared with the condition in which they simulated an incongruent expression to a negative picture. The latter condition would have prevented targets displaying a convincing simulation of an incongruent expression, thus providing clues to perceivers that their behaviour is simulated.

In summary, this method of asking perceivers to make retrodictive inferences has overcome many limitations of past studies. But some limitations

remain, allowing scope for even further methodological improvements in the future.

8.5 DIRECTIONS FOR FUTURE RESEARCH

When people make a retrodictive inference, it is assumed that their performance would be different than people who merely classified their responses. However, there was no substantial difference in the performance of retrodicting and classifying perceivers, raising the possibility that both groups were engaging in a similar processing strategy. Nevertheless, it could be possible that the two groups used different strategies but that the kind of measures used in this research were not sensitive enough to detect different strategies. A possible way to investigate these differences would be to use neuroimaging techniques such as electroencephalogram (EEG) or functional magnetic resonance imaging (fMRI) to measure brain activity. It is expected that classifying perceivers would show stronger activity in the face processing event-related potential (ERP) components such as the early components involved in structural encoding (N170) (Bentin et al., 1996; Kloth et al., 2010; Maurer et al., 2008), P200 (Paulmann and Pell, 2009) and mid-latency components between 300 and 400 ms (Adolphs, 2002; Cuthbert et al., 2000; Eimer, 2000). In contrast, retrodicting perceivers might show stronger signals in components known to be activated when they are making inner state inferences, as in false belief tasks. Such tasks evoke frontal and parietal slow wave activities (Sabbagh and Taylor, 2000; Liu et al., 2004, 2009). Other than ERP components, it is expected that there will be stronger brain activation in areas that are involved in making inference of inner states. The right temporo-parietal

junction is activated during the prediction of actions from false belief tasks (Saxe and Kanwisher, 2003; Vogeley et al. 2001) while the medial prefrontal cortex is activated during the processing of social and emotional information (Mitchell et al., 2005) and inner state terms (Gallagher et al., 2000; Saxe and Kanwisher, 2003; Young et al., 2010). In a study by Mier (2010), activation in the inferior frontal gyrus, superior temporal sulcus, temporal pole and amygdala was more pronounced for intention recognition as compared to emotion recognition. Granted, brain activation during retrodiction tasks might be different from prediction tasks. However, these neuroimaging techniques could be utilised as a complementary technique to uncover the different processing strategies used by the two perceiver groups.

Other than using neuroimaging techniques, it would be interesting to investigate if people with alexithymia can make retrodictive inferences. Alexithymia is often associated with impaired verbal and nonverbal recognition of emotional stimuli and difficulty in putting emotions into words (Lane et al., 1996). From the recognition of morphed and posed facial expressions, Cook et al. (2013) discovered that people with alexithymia were unable to interpret the emotional content of the face. Alexithymia traits, too, predicted performance on emotion recognition from the eyes region but not when they were asked to answer some questions pertaining to the actors' inner states from short films, as assessed by the Movie Assessment for Social Cognition (MASC) test (Oakley et al., 2016). If retrodictive inferences required the identification of inner states such as emotions, as suggested by the simulation account of mentalizing (Gallese and Goldman, 1998; Teoh et al., 2017), it is expected that perceivers with alexithymia would not perform accurately on these tasks. This is because retrodiction might involve a

form of mental simulation which involves experiencing and identifying the inner states that targets are experiencing. If perceivers with alexithymia cannot identify their own inner states, then presumably they would not be able to carry out the process of mental simulation.

If, however, perceivers with alexithymia do make accurate retrodictive inferences, then this suggests that, while individuals with alexithymia are impaired at recognising emotions, they are not impaired at inferring an external event that happened in the world. Since the recognition of emotions is impaired in individuals with alexithymia, they might be engaging in an alternative strategy other than inferring emotions of targets. There are two explanations as to how individuals with alexithymia might be able to make accurate retrodictive inferences despite being impaired at recognising emotions. Firstly, perceivers with alexithymia might be matching the specific facial configuration with the corresponding situation according to a system of behavioural rules, bypassing any inference of inner states. This would also indicate that retrodictive inferences does not necessarily require the inference of inner states. Secondly, as Oakley et al. (2016) discovered that individuals with alexithymia could infer others' inner states despite being impaired at recognising emotions, they might be inferring other types of inner states apart from emotions which will, in turn, guide their retrodictive inferences. Thus, it would be interesting to investigate if individuals with alexithymia can make accurate retrodictive inferences in order to shed light onto the processes underlying retrodiction.

Furthermore, in real-life, people make inferences about others from facial expressions, body language and voice. Although individuals can make accurate

retrodictive inferences from spontaneous and simulated congruent facial expressions alone, individuals are unable to make these inferences from simulated incongruent facial expressions. However, within the deception literature, body language has been shown to play an important role in detecting deception. In Ekman and Friesen (1974)'s study, targets were asked to be honest in one interview by frankly describing their feelings about a film, or to deceive others by concealing their negative affect by simulating pleasant feelings. When perceivers viewed a sample of targets' behaviour from honest interviews, they made more accurate judgements from the body than the face. This could be because targets placed too much emphasis on concealing their facial expressions as they believed that their expressions would betray their true inner states. Thus, it would be interesting to investigate if perceivers could make more accurate inferences from incongruent expressions if targets' body language was visible, in addition to facial expressions.

The findings in this thesis can be extended in a variety of ways. Future studies should utilise neuroimaging techniques to identify the underlying mechanisms as these techniques are sensitive to different processing mechanisms which are not otherwise apparent in behaviour and eye movements. Additionally, it would also be interesting to test the retrodictive ability of individuals in alexithymia to further identify the processes involved in retrodiction. Lastly, in addition to facial expressions, bodily and voice cues should also be included to further improve the ecological validity of the stimuli.

8.6 CONCLUSION

The studies from this thesis provided an insight into people's ability to make retrodictive inferences from spontaneous and simulated expressions. Under natural circumstances when one is behaving spontaneously, people are above chance level when making retrodictive inferences as to what others were viewing, regardless of whether it was the emotional expression of another person or the emotional content of a picture. Similar to previous retrodiction studies (Cassidy et al., 2014; Pillai et al., 2012; Teoh et al., 2017), these findings imply that people's minds are signalled in their behaviour, a signal which provides clues for others to make inferences about their inner states and the antecedent event.

The research reported in this thesis also extends past studies by investigating people's ability to make retrodictive inferences from simulated expressions. While people are able to see through others' behaviour to identify if they are simulating a congruent or incongruent expression, this was only the case if targets were experiencing something negative. Despite being able to identify if targets were behaving in an incongruent manner, this incongruent behaviour still obstructed perceivers from accessing targets' minds in such a way that allows them to infer the true inner state of targets and the event that they were responding to. As such, if targets are inclined to fool perceivers then perceivers cannot see through the target's behaviour to infer their true inner states and what had happened to them.

When people retrodict from a person's behaviour to the event that caused the behaviour, it is uncertain as to whether people are genuinely inferring the inner states of others. Even though retrodicting perceivers performed similarly to

classifying perceivers, it does not necessarily follow that both groups of perceivers do not make inferences about inner states. There is also a possibility that both groups of perceivers might be engaging in different strategies but these differences are not apparent in this study due to the insensitivity of the measures used. Thus, no firm conclusion can be made about the processing strategies involved in the tasks.

Nevertheless, the ability to make retrodictive inferences about others is adaptive because it permits us to make sense of others' behaviour and to use this behaviour as an effective tool for knowing about external events that happened in the world. In our daily lives, we often observe people's behaviour and attempt to determine the cause and motivation behind their behaviour; this ability to make retrodictive inferences is thus an essential skill that confers a level of adaptiveness in the social arena. The findings have highlighted the value of using subtle, naturalistic and ecologically valid behaviour to investigate social perception in an experimental procedure. While this experimental paradigm of retrodictive inference represents a leap forward from past research, future studies should include body language and voice to further inform us about people's ability to make these inferences under real-life circumstances. Additionally, future studies should continue to aim to uncover the processing strategy underlying retrodictive inferences.

REFERENCES

Adolphs, R. (1999). Social cognition and the human brain. *Trends in cognitive sciences*, *3*(12), 469-479.

Adolphs, R. (2002). Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews*, *1*(1), 21-62.

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*(6507), 669-672.

Ambadar, Z., Schooler, J. W., & Cohn, J. F. (2005). Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions. *Psychological science*, *16*(5), 403-410.

Anderson, D. E., DePaulo, B. M., & Ansfield, M. E. (2002). The development of deception detection skill: A longitudinal study of same-sex friends. *Personality and Social Psychology Bulletin*, *28*(4), 536-545.

Andrews, K. (2005). Chimpanzee theory of mind: looking in all the wrong places?. *Mind & Language*, *20*(5), 521-536.

Argyle, M. (1978). The laws of looking. *Human Nature*, *1*, 32.

Baddeley, A., & Woodhead, M. (1981). Techniques for improving eyewitness identification skills. Paper presented at the *SSRC law and Psychology Conference*, Trinity College, Oxford, April.

Baron-Cohen, S. (1996). Reading the mind in the face: A cross-cultural and developmental study. *Visual Cognition*, *3*(1), 39-60.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, A. T. (1997). Is there a" language of the eyes"? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition*, *4*(3), 311-331.

Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child development*, *60*(4), 946-964.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, *8*(6), 551-565.

Biehl, M., Matsumoto, D., Ekman, P., & Hearn, V. (1997). Matsumoto and Ekman's Japanese and Caucasian Facial Expressions of Emotion (JACFEE): Reliability data and cross-national differences*. Journal of Nonverbal Behavior, 21*(1), 3–21.

Blair, I. V., Judd, C. M., Sadler, M. S., & Jenkins, C. (2002). The role of Afrocentric features in person perception: judging by features and categories. *Journal of Personality and Social Psychology*, *83*(1), 5-25.

Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of Nonverbal behavior*, *23*(1), 5-41.

Blais, C., Roy, C., Fiset, D., Arguin, M., & Gosselin, F. (2012). The eyes are not the window to basic emotions. *Neuropsychologia*, *50*(12), 2830-2838.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and social psychology Review*, *10*(3), 214-234.

Bond, C. F., Omar, A., Pitre, U., Lashley, B. R., Skaggs, L. M., & Kirk, C. T. (1992). Fishy-looking liars: deception judgment from expectancy violation. *Journal of personality and social psychology*, *63*(6), 969-977.

Boucher, J. D., & Carlson, G. E. (1980). Recognition of facial expression in three cultures. *Journal of Cross-Cultural Psychology*, *11*(3), 263-280.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305-327.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of*

Experimental Psychology, *66*(8), 1467-1485.

Burton, A. M., Bruce, V., & Dench, N. (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, *22*(2), 153-176.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943-958.

Burton, A. M., Kramer, R. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, *40*(1), 202-223.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision research*, *41*(9), 1179-1208.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human perception and performance*, *26*(2), 527-551.

Calvo, M. G., & Lundqvist, D. (2008). Facial expressions of emotion (KDEF): Identification under different display-duration conditions. *Behaviour research methods*, *40*(1), 109-115.

Calvo, M. G., Gutiérrez-García, A., Avero, P., & Lundqvist, D. (2013). Attentional mechanisms in judging genuine and fake smiles: Eye-movement patterns. *Emotion*, *13*(4), 792-802.

Campbell, R., Elgar, K., Kuntsi, J., Akers, R., Terstegge, J., Coleman, M., & Skuse, D. (2002). The classification of 'fear' from faces is associated with face recognition skill in women. *Neuropsychologia*, *40*(6), 575-584.

Cassidy, S., Mitchell, P., Chapman, P., & Ropar, D. (2015). Processing of spontaneous emotional responses in adolescents and adults with autism spectrum disorders: effect of stimulus type. *Autism Research*, *8*(5), 534-544.

Cassidy, S., Ropar, D., Mitchell, P., & Chapman, P. (2014). Can adults with autism spectrum disorders infer what happened to someone from their emotional response?. *Autism Research*, *7*(1), 112-123.

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. *Memory & cognition*, *26*(4), 780-790.

Cole, T. (2001). Lying to the one you love: The use of deception in romantic relationships. *Journal of Social and Personal Relationships*, *18*(1), 107-129.

Collignon, O., Girard, S., Gosselin, F., Saint-Amour, D., Lepore, F., & Lassonde, M. (2011). Women process multisensory emotion expressions more efficiently than men. *Neuropsychologia, 48*(1), 220–225.

Cook, R., Brewer, R., Shah, P., & Bird, G. (2013). Alexithymia, not autism, predicts poor recognition of emotional facial expressions. *Psychological Science*, *24*(5), 723-732.

Cottrell, G. W., Dailey, M. N., Padgett, C., & Adolphs, R. (2001). Is all face processing holistic? The view from UCSD. In Wenger, M.J. and Townsend, J.T (Eds.), *Computational, Geometric, and Process Perspectives on Facial Cognition* (pp. 347– 396). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cottrell, G. W., Padgett, C., & Adolphs, R. (1998). Modelling megamix: A computational model of categorical perception of facial expressions. *Journal of Cognitive Neuroscience, 10,* 36.

Cunningham, D. W., Kleiner, M., Bülthoff, H. H., & Wallraven, C. (2004). The components of conversational facial expressions. In *Proceedings of the 1st Symposium on Applied perception in graphics and visualization* (pp. 143-150). New York, NY, USA: ACM.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological psychology*, *52*(2), 95-111.

Davies, G., Ellis, H., & Shepherd, J. (1977). Cue saliency in faces as assessed by the 'Photofit' technique. *Perception*, *6*(3), 263-269.

Darwin, C. (1965). *The expression of emotion in animals and man.* Chicago: University of Chicago Press. (Original work published 1872)

de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of personality and social psychology*, *106*(1), 1-16.

DePaulo, B. M. (1994). Spotting lies: Can humans learn to do better?. *Current directions in psychological science*, *3*(3), 83-86.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of personality and social psychology*, *74*(1), 63.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of personality and social psychology*, *70*(5), 979-995.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, *129*(1), 74-118.

DePaulo, B. M. & Morris, W. L. (2004). Discerning lies from truths: Behavioural cues to deception and the indirect pathway of intuition. In Granhag, P. A. & Strömwall, L. A. *The detection of deception in forensic contexts (*pp. 15–40). Cambridge, England: Cambridge University Press.

DePaulo, B. M., & Pfeifer, R. L. (1986). On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology*, *16*(3), 249-267.

DePaulo, B. M., Stone, J. I., & Lassiter, G. D. (1985). Telling ingratiating lies: Effects of target sex and target attractiveness on verbal and nonverbal deceptive success. *Journal of Personality and Social Psychology*, *48*(5), 1191-1203.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological science*, *11*(1), 86-89.

Eimer, M. (2000). Attentional modulations of event-related brain potentials sensitive to faces. *Cognitive Neuropsychology*, *17*(1-3), 103-116.

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion*, *11*(4), 860-865.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.

Ekman, P. (2003). Darwin, deception, and facial expression. *Annals of the New York Academy of Sciences*, *1000*(1), 205-221.

Ekman, P. (2004). Emotional and conversational nonverbal signals. In Larrazabal, J.M. & Miranda, L.A. (Eds.) *Language, knowledge, and representation* (pp. 39-50). Dordrecht: Springer.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124-129.

Ekman, P., & Friesen, W. V. (1974). Detecting deception from the body or face. *Journal of personality and Social Psychology*, *29*(3), 288-298.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues.* Englewood Cliffs, NJ: Prentice-Hall.

Ekman, P., & Friesen, W.V. (1976). *Pictures of facial affect*. Consulting Psychologists Press. CA: Palo Alto.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of nonverbal behavior*, *6*(4), 238-252.

Ekman, P., & Friesen, W. V. (1986). A new pan-cultural facial expression of

emotion. *Motivation and emotion*, *10*(2), 159-168.

Ekman, P., & Heider, K. G. (1988). The universality of a contempt expression: A replication. *Motivation and Emotion*, *12*(3), 303-308.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar?. *American psychologist*, *46*(9), 913-920.

Ekman, P., & O'Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: the utility of voluntary and involuntary behavior in detecting deception. *Behavioral sciences & the law*, *24*(5), 673-686.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the Human Face: Guidelines for Research and an Integration of Findings: Guidelines for Research and an Integration of Findings*. New York: Pergamon.

Ekman, P., Friesen, W. V., & O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology*, *54*(3), 414-420.

Ekman, P., O'Sullivan, M., Friesen, W. V., & Scherer, K. R. (1991). Invited article: Face, voice, and body in detecting deceit. *Journal of nonverbal behavior*, *15*(2), 125-135.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, *164*(3875), 86-88.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychological bulletin*, *128*(2), 203-235.

Elfenbein, H. A., & Ambady, N. (2003). When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *Journal of personality and social psychology*, *85*(2), 276-290.

Elfenbein, H. A., Beaupré, M., Lévesque, M., & Hess, U. (2007). Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion*, *7*(1), 131-146.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception*, *8*(4), 431-439.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of cognitive neuroscience*, *13*(2), 171-180.

Flavell, J., Green F.L., & Flavell, F.R. (1993). Children's understanding of the stream of consciousness. *Child Development*, *64*(2), 387–398.

Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across

different types of high-stake lies. *Journal of personality and social psychology*, *72*(6), 1429-1439.

Fraser, I. H., Craig, G. L., & Parker, D. M. (1990). Reaction time measures of feature saliency in schematic faces. *Perception*, *19*(5), 661-673.

Freeth, M., Chapman, P., Ropar, D., & Mitchell, P. (2010). Do gaze cues in complex scenes capture and direct the attention of high functioning adolescents with ASD? Evidence from eye-tracking. *Journal of autism and developmental disorders*, *40*(5), 534-547.

Frijda, N. H. (1986). *The emotions: Studies in emotion and social interaction.* Paris: Maison de Sciences de l'Homme.

Gagnon, M., Gosselin, P., Hudon-ven Der Buhs, I., Larocque, K., & Milliard, K. (2010). Children's recognition and discrimination of fear and disgust facial expressions. *Journal of Nonverbal Behavior, 34*(1), 27-42.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11-21.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17*(2), 535-543.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Gao, X., & Maurer, D. (2009). Influence of intensity on children's sensitivity to happy, sad, and fearful facial expressions. *Journal of experimental child psychology*, *102*(4), 503-521.

Gao, X., & Maurer, D. (2010). A happy story: Developmental changes in children's sensitivity to facial expressions of varying intensities. *Journal of Experimental Child Psychology*, *107*(2), 67-86.

Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture. *Emotion*, *14*(2), 251-262.

Gepner, B., Deruelle, C., & Grynfeltt, S. (2001). Motion and emotion: A novel approach to the study of face processing by young autistic children. *Journal of autism and developmental disorders*, *31*(1), 37-45.

Goldberg, J. H. (2000). Eye movement-based interface evaluation: What can and cannot be assessed?. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(37), 625-628. CA: SAGE Publications.

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631-645.

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition*, *94*(3), 193-213.

Gosselin, P., & Kirouac, G. (1995). Decoding facial emotional prototypes. *Canadian Journal of Experimental Psychology= Revue canadienne de psychologie experimentale, 49*(3), 313-329.

Gosselin, P., & Simard, J. (1999). Children's knowledge of facial expressions of emotions: Distinguishing fear and surprise. *The Journal of genetic psychology, 160*(2), 181-193.

Granhag, P. A., & Hartwig, M. (2008). A new theoretical perspective on deception detection: On the psychology of instrumental mind-reading. *Psychology, Crime & Law*, *14*(3), 189-200.

Green, M. (2007). *Self-expression*. Oxford: Oxford University Press

Hanawalt, N. (1944). The role of the upper and lower parts of the face as the basis for judging facial expressions: II. In posed expressions and “candid camera” pictures. *Journal of General Psychology, 31,* 23–36.

Hareli, S., Kafetsios, K., & Hess, U. (2015). A cross-cultural study on emotion expression and the learning of social norms. *Frontiers in psychology*, *6*.

Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American Journal on mental retardation*, *104*(3), 270-278.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & cognition*, *33*(1), 98-106.

Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: Behavioural and neurological perspectives. *Journal of Child Psychology and Psychiatry*, *45*(7), 1185-1198.

Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia*, *47*(4), 1004-1012.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding

accuracy. *International journal of psychophysiology*, *40*(2), 129-141.

Hess, U., Blaison, C., & Kafetsios, K. (2016). Judging facial emotion expressions in context: The influence of culture and self-construal orientation. *Journal of Nonverbal Behavior*, *40*(1), 55-64.

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, *20*(5), 369-385.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131-143.

Hill, H., & Johnston, A. (2001). Categorizing sex and identity from the biological motion of faces. *Current biology*, *11*(11), 880-885.

Ickes, W. J. (1997). *Empathic accuracy*. Guilford Press.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*(4), 730.

Izard, C. E. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Izard, C. E. (1977). *Human emotions*. New York: Plenum

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural

confusions show that facial expressions are not universal. *Current Biology*, *19*(18), 1543-1548.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, *109*(19), 7241-7244.

Janik, S. W., Wellens, A. R., Goldberg, M. L., & Dell'Osso, L. F. (1978). Eyes as the center of focus in the visual examination of human faces. *Perceptual and motor skills*, *47*(3), 857-858.

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition*, *121*(3), 313-323.

Jensen, L. A., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2004). The right to do wrong: Lying to parents among adolescents and emerging adults. *Journal of Youth and Adolescence*, *33*(2), 101-112.

Johnston, P. J., Katsikitis, M., & Carr, V. J. (2001). A generalised deficit can account for problems in facial emotion recognition in schizophrenia. *Biological psychology*, *58*(3), 203-227.

Kaliouby, E. R., & Robinson, P. (2005). Real-time inference of complex mental states from facial expressions and head gestures. In Kisacanin, B.,

Pavlovic, V. & Huang, T.S. (Eds.) *Real-time vision for human-computer interaction* (pp. 181-200). Heidelberg: Springer.

Kamachi, M., Bruce, V., Mukaida, S., Gyoba, J., Yoshikawa, S., & Akamatsu, S. (2013). Dynamic properties influence the perception of facial expressions. Perception, 42(11), 1266-1278.

Kirchner, J. C., Hatri, A., Heekeren, H. R., & Dziobek, I. (2011). Autistic symptomatology, face processing abilities, and eye fixation patterns. *Journal of autism and developmental disorders*, *41*(2), 158-167.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, *59*(9), 809-816.

Kloth, N., Schweinberger, S. R., & Kovács, G. (2010). Neural correlates of generic versus gender-specific face adaptation. *Journal of Cognitive Neuroscience*, *22*(10), 2345-2356.

Köhnken, G. (1987). Training police officers to detect deceptive eyewitness statements: Does it work? *Social Behaviour, 2*(1), 1-17.

Kraut, R. E., & Poe, D. B. (1980). Behavioral roots of person perception: The

deception judgments of customs inspectors and laymen. *Journal of Personality and Social Psychology*, *39*(5), 784-798.

Krumhuber, E. G., Kappas, A., & Manstead, A. S. (2013). Effects of dynamic aspects of facial expressions: a review. *Emotion Review*, *5*(1), 41-46.

Laird, J. D., Alibozak, T., Davainis, D., Deignan, K., Fontanella, K., Hong, J., & Pacheco, C. (1994). Individual differences in the effects of spontaneous mimicry on emotional contagion. *Motivation and Emotion*, *18*(3), 231-247.

Lander, K., & Bruce, V. (2000). Recognizing famous faces: Exploring the benefits of facial motion. *Ecological Psychology*, *12*(4), 259-272.

Lane, R. D., Lee, S., Reidel, R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic medicine*, *58*(3), 203-210.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings.

Lawrence, K., Campbell, R., & Skuse, D. (2015). Age, gender, and puberty influence the development of facial emotion recognition. *Frontiers in Psychology*, *6*.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764-766.

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: an ERP study of theory of mind. *NeuroReport*, *15*(6), 991-995.

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2009). Neural correlates of children's theory of mind development. *Child development*, *80*(2), 318-326.

Lundqvist, L. O., & Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology*, *9*(3), 203-211.

Lundqvist, D., Flykt, A., & Öhman, A. (1998). *The Karolinska directed emotional faces (KDEF).* CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior research methods*, *47*(4), 1122-1135.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group

data: estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, *98*(1), 185.

Mandai, M. K., &. Palchoudhury, S. (1985). Perceptual skill in decoding facial affect. *Perceptual &: Motor Skills, 60*(1), 96-98.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & Van de Veerdonk, E. (2008). Placing the face in context: cultural differences in the perception of facial emotion. *Journal of personality and social psychology*, *94*(3), 365-381.

Matsumoto, D. (1992). More evidence for the universality of a contempt expression. *Motivation and Emotion*, *16*(4), 363-368.

Matsumoto, D., Yoo, S. H., & Nakagawa, S. (2008). Culture, emotion regulation, and adjustment. *Journal of personality and social psychology*, *94*(6), 925-937.

Maurer, U., Rossion, B., & McCandliss, B. D. (2008). Category specificity in early perception: face and word N170 responses differ in both lateralization and habituation properties. *Frontiers in human neuroscience*, *2*.

McAndrew, F. T. (1986). A cross-cultural study of recognition thresholds for

facial expressions of emotion. *Journal of Cross-Cultural Psychology*, *17*(2), 211-224.

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, *47*(6), 1028-1039.

Mitchell, J. P., Banaji, M. R., & Macrae, C. N. (2005). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*, *28*(4), 757-762.

Montirosso, R., Peverelli, M., Frigerio, E., Crespi, M., & Borgatti, R. (2010). The Development of Dynamic Facial Expression Recognition at Different Intensities in 4-to 18-Year-Olds. *Social Development*, *19*(1), 71-92.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, *9*(5), 555-604.

Motley, M. T., & Camden, C. T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of*

Communication (includes Communication Reports), *52*(1), 1-22.

Nelson, N. L., & Russell, J. A. (2013). Universality revisited. *Emotion Review*, *5*(1), 8-15.

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, Å. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression. *Cognition & Emotion*, *15*(6), 853-864.

North, M. S., Todorov, A., & Osherson, D. N. (2010). Inferring the preferences of others from spontaneous, low-emotional facial expressions. *Journal of Experimental Social Psychology*, *46*(6), 1109-1113.

North, M. S., Todorov, A., & Osherson, D. N. (2012). Accuracy of inferring self- and other-preferences from spontaneous facial expressions. *Journal of Nonverbal Behavior*, *36*(4), 227-233.

Oakley, B. F., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of abnormal psychology*, *125*(6), 818-823.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: a threat advantage with schematic stimuli. *Journal of personality and*

social psychology, *80*(3), 381-396.

Orgeta, V. (2010). Effects of age and task difficulty on recognition of facial affect. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *65*(3), 323-327.

Padgett, C., & Cottrell, G. W. (1998). A simple neural network models categorical perception of facial expressions. In *Proceedings of the twentieth annual cognitive science conference*, 806-807.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(4), 406-417.

Paulmann, S., & Pell, M. D. (2009). Facial expression decoding as a function of emotional meaning status: ERP evidence. *NeuroReport*, *20*(18), 1603-1608.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1), 8-13.

Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions

in a foreign language. *Journal of Nonverbal Behavior*, *33*(2), 107-120.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep?. *Science*, *308*(5719), 214-216.

Pillai, D., Sheppard, E., & Mitchell, P. (2012). Can people guess what happened to others from their reactions?. *PLoS One*, *7*(11), e49859.

Pillai, D., Sheppard, E., Ropar, D., Marsh, L., Pearson, A., & Mitchell, P. (2014). Using other minds as a window onto the world: guessing what happened from clues in behaviour. *Journal of autism and developmental disorders*, *44*(10), 2430-2439.

Porter, S., & Ten Brinke, L. (2008). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological science*, *19*(5), 508-514.

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human?. *Trends in cognitive sciences*, *7*(4), 157-160.

Povinelli, D. J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind & Language*, *19*(1), 1-28.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526.

Rai, R., & Mitchell, P. (2004). Five-year-old children's difficulty with false belief when the sought entity is a person. *Journal of experimental child psychology*, *89*(2), 112-126.

Ratcliffe, M. J. (2007). Rethinking commonsense psychology: A critique of folk psychology, *theory of mind and simulation.* UK: Palgrave Macmillan.

Reinhard, M. A., Scharmach, M., & Stahlberg, D. (2013). Too exhausted to see the truth: Ego depletion and the ability to detect deception. *British Journal of Social Psychology*, *52*(4), 618-630.

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *The Quarterly Journal of Experimental Psychology*, *70*(5), 897-905.

Robinson, E. J., & Mitchell, P. (1995). Masking of children's early understanding of the representational mind: Backwards explanation versus prediction. *Child Development*, 1022-1039.

Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of facial expression recognition. *Developmental Science*, *18*(6), 926-939.

Rotter, N. G., & Rotter, G. S. (1988). Sex differences in the encoding and

decoding of negative facial emotions. *Journal of Nonverbal Behavior*, *12*(2), 139-148.

Russell, J. A. (1993). Forced-choice response format in the study of facial expression. *Motivation and Emotion*, *17*(1), 41-51.

Russell, J. A., Bachorowski, J. A., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual review of psychology*, *54*(1), 329-349.

Sabbagh, M. A., & Taylor, M. (2000). Neural correlates of theory-of-mind reasoning: an event-related potential study. *Psychological Science*, *11*(1), 46-50.

Sato, W., & Yoshikawa, S. (2007). Spontaneous facial mimicry in response to dynamic facial expressions. *Cognition*, *104*(1), 1-18.

Sato, W., Kochiyama, T., Yoshikawa, S., Naito, E., & Matsumura, M. (2004). Enhanced neural activity in response to dynamic facial expressions of emotion: an fMRI study. *Cognitive Brain Research*, *20*(1), 81-91.

Sauter, D. A., & Fischer, A. H. (2017). Can perceivers recognise emotions from spontaneous expressions?. *Cognition and Emotion*, 1-12.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of

mind". *Neuroimage*, *19*(4), 1835-1842.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in cognitive sciences*, *4*(8), 299-309.

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of vision*, *14*(13), 1-16.

Serota, K. B., Levine, T. R., & Boster, F. J. (2010). The prevalence of lying in America: Three studies of self-reported lies. *Human Communication Research*, *36*(1), 2-25.

Serrano, J. M., Iglesias, J., & Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprised in 4-to 6-month-old infants. *Developmental Psychobiology*, *25*(6), 411-425.

Smortchkova, J. (2017). Seeing emotions without mindreading them. *Phenomenology and the Cognitive Sciences*, *16*(3), 525-543.

Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signalling: Its effect on the visual cliff behaviour of 1-year-olds. *Developmental psychology*, *21*(1), 195-2000.

Sporer, S. L., & Schwandt, B. (2006). Paraverbal indicators of deception: A meta-

analytic synthesis. *Applied Cognitive Psychology*, *20*(4), 421-446.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law, 13*(1), 1-34.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, *46*(2), 225-245.

Tcherkassof, A., Bollon, T., Dubois, M., Pansu, P., & Adam, J. M. (2007). Facial expressions of emotions: A methodological contribution to the study of spontaneous and dynamic emotional faces. *European Journal of Social Psychology*, *37*(6), 1325-1345.

ten Brinke, L., Stimson, D., & Carney, D. R. (2014). Some evidence for unconscious lie detection. *Psychological Science*, *25*(5), 1098-1105.

ten Brinke, L., Vohs, K. D., & Carney, D. R. (2016). Can ordinary people detect deception after all?. *Trends in cognitive sciences*, *20*(8), 579-588.

Teoh, Y., Wallis, E., Stephen, I. D., & Mitchell, P. (2017). Seeing the world through others' minds: Inferring social context from behaviour. *Cognition*, *159*, 48-60.

Thayer, J. F., & Johnson, B. H. (2000). Sex differences in judgment of facial affect:

A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology, 41*, 243–246.

The Psychological Image Collection at Stirling (PICS) (n.d.). University of Stirling Psychology Department. Retrieved January, 2013 from http://pics.psych.stir.ac.uk/.

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, *28*(7), 1164-1195.

Thornton, I. M., & Kourtzi, Z. (2002). A matching advantage for dynamic human faces. *Perception*, *31*(1), 113-132.

Tomkins, S. S. (1962). *Affect, imagery, consciousness. Vol. 1. The positive affects*. New York: Springer.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Research*, *168*(3), 242-249.

Uttal, W. R. (1999). *The war between mentalism and behaviorism: On the accessibility of mental processes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Vogeley, K., et al. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage*, *14*(1), 170-181.

Vrij, A., Edward, K., & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and social psychology bulletin*, *27*(7), 899-909.

Vrij, A., Edward, K., Roberts, K. P., & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal behavior*, *24*(4), 239-263.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, *5*(1-2), 39-43.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of nonverbal behavior*, *17*(1), 3-28.

Wagner, H. L., MacDonald, C. J., & Manstead, A. S. (1986). Communication of individual emotions by spontaneous facial expressions. *Journal of Personality and Social Psychology*, *50*(4), 737.

Wang, K., Hoosain, R., Lee, T. M., Meng, Y., Fu, J., & Yang, R. (2006). Perception of six basic emotional facial expressions by the Chinese. *Journal of Cross-*

Cultural Psychology, *37*(6), 623-629.

Warren, G., Schertler, E., & Bull, P. (2009). Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior*, *33*(1), 59-69.

Weiner, B. (2006). *Social motivation, justice, and the moral emotions: An attributional approach*, Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child development*, *72*(3), 655-684.

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological review*, *118*(2), 357-378.

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review*, *5*(1), 72-77.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Wood, A., Rychlowska, M., Korb, S., & Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression

recognition. *Trends in cognitive sciences*, *20*(3), 227-240.

Woodworth, R. S. (1938). *Experimental Psychology.* New York: Holt.

Yarbus, A. L. (1967). *Eye movements and vision*. US: Springer.

Yoshikawa, S., & Sato, W. (2006). Enhanced perceptual, emotional, and motor processing in response to dynamic facial expressions of emotion. *Japanese Psychological Research*, *48*(3), 213-222.

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research*, *48*(2), 63-68.

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, *107*(15), 6753-6758.

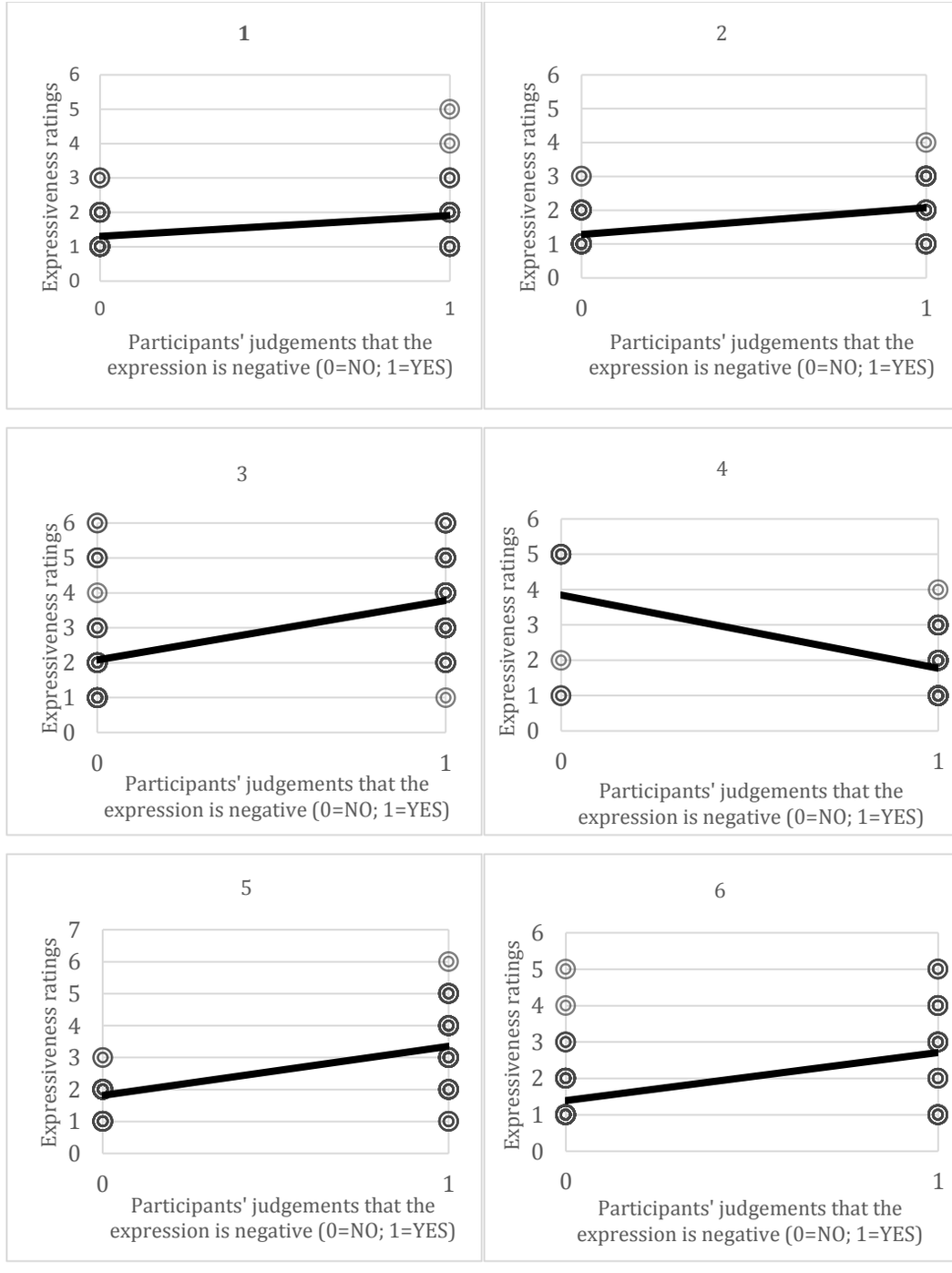
Zahavi, D. & Gallagher, S. (2008) *The Phenomenological Mind*. New York: Routledge.

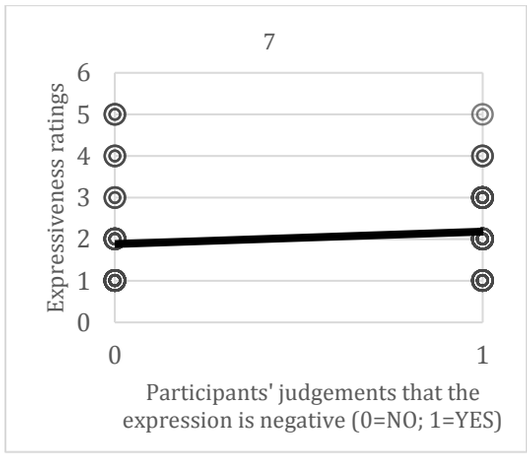
Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion*, *9*(4), 478-487.

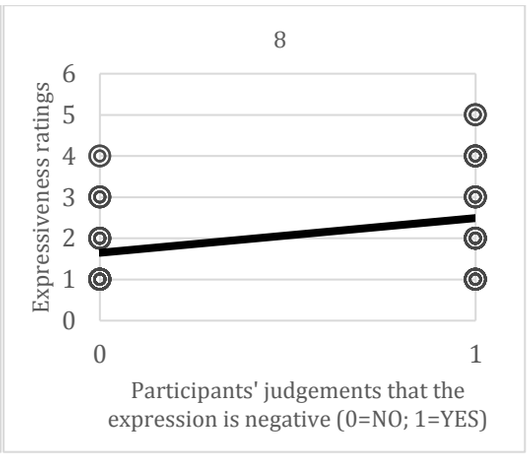
Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior*, *6*(2), 105-114.

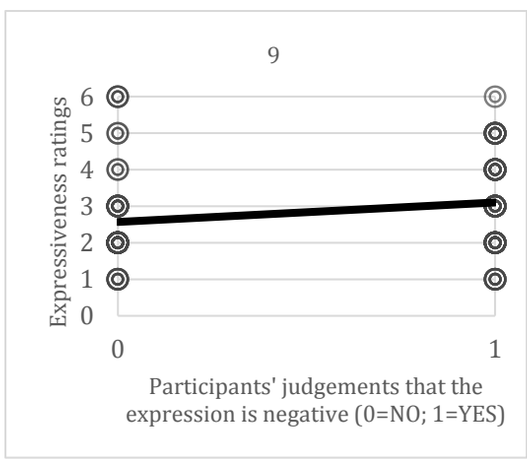
SUPPLEMENTARY MATERIALS

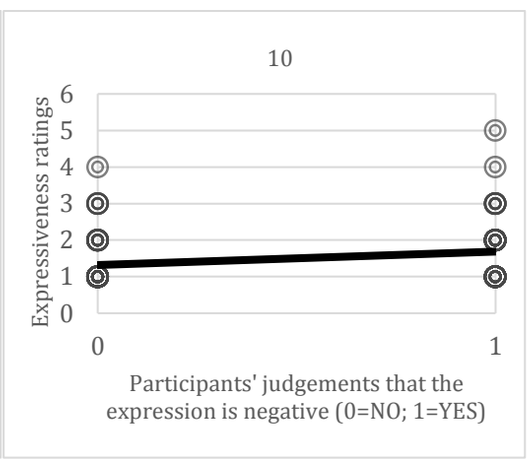
A. The relationship between expressiveness ratings and participants' responses that targets' expressions was negative, for all judges

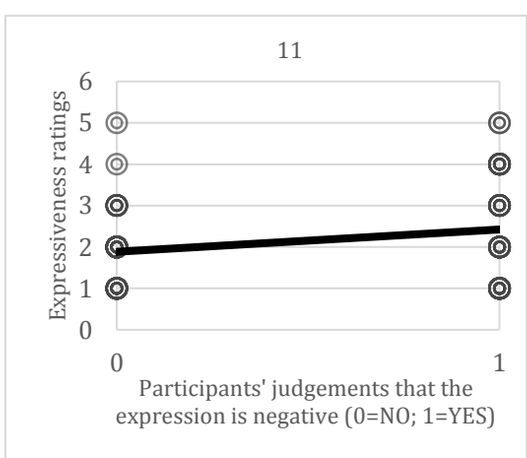


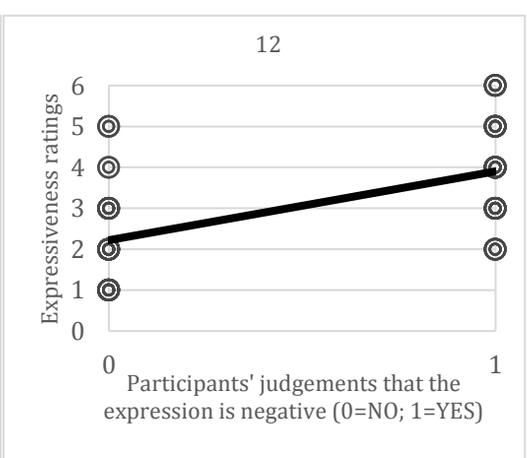

7
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

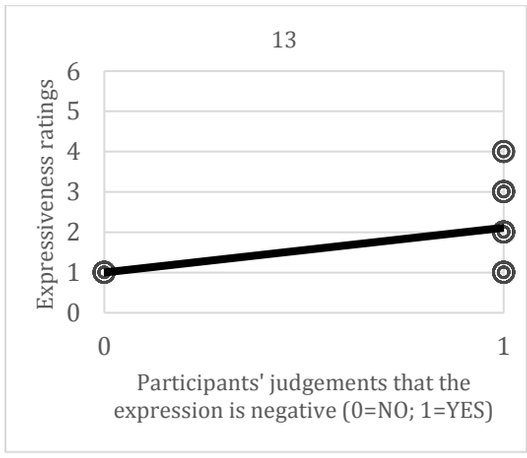

8
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

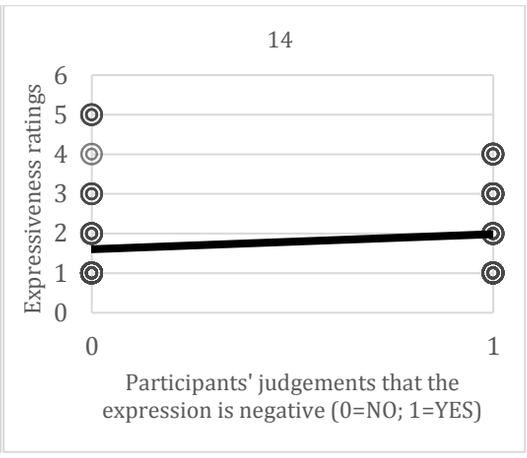

9
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

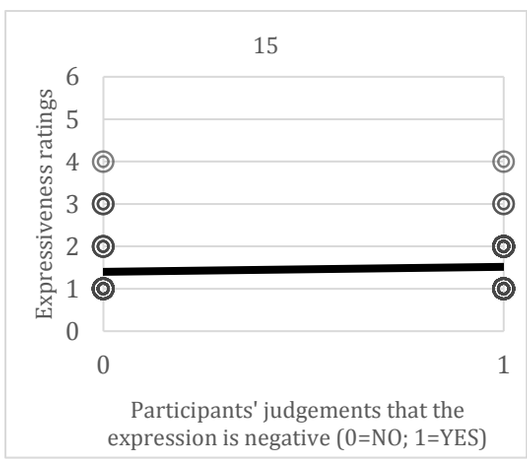

10
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

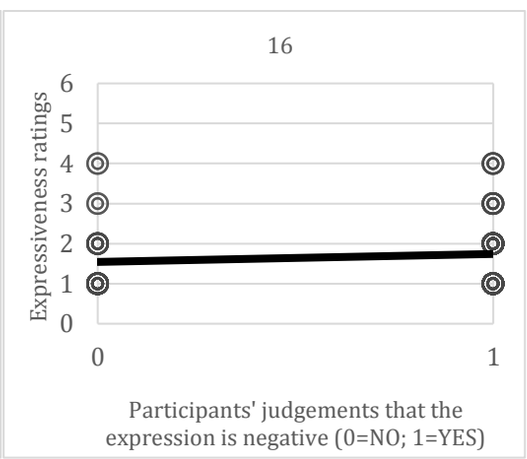

11
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

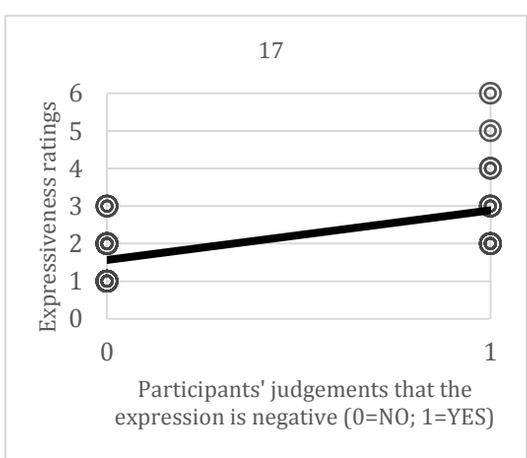

12
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

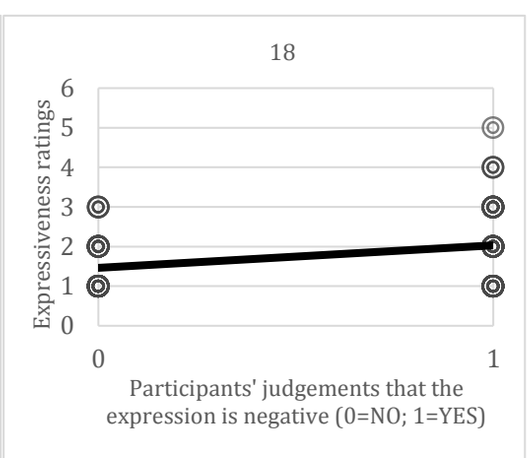

13
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

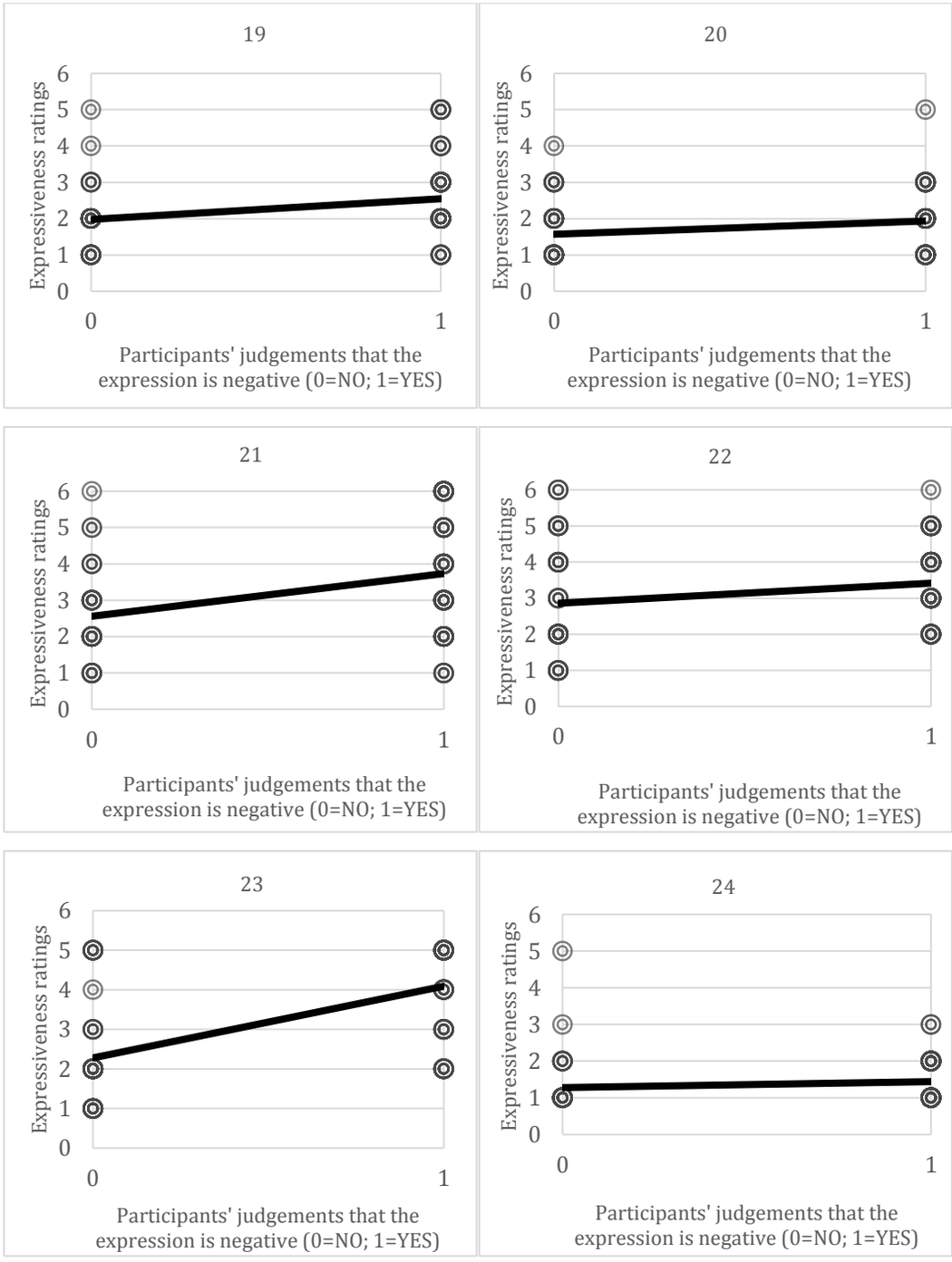

14
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

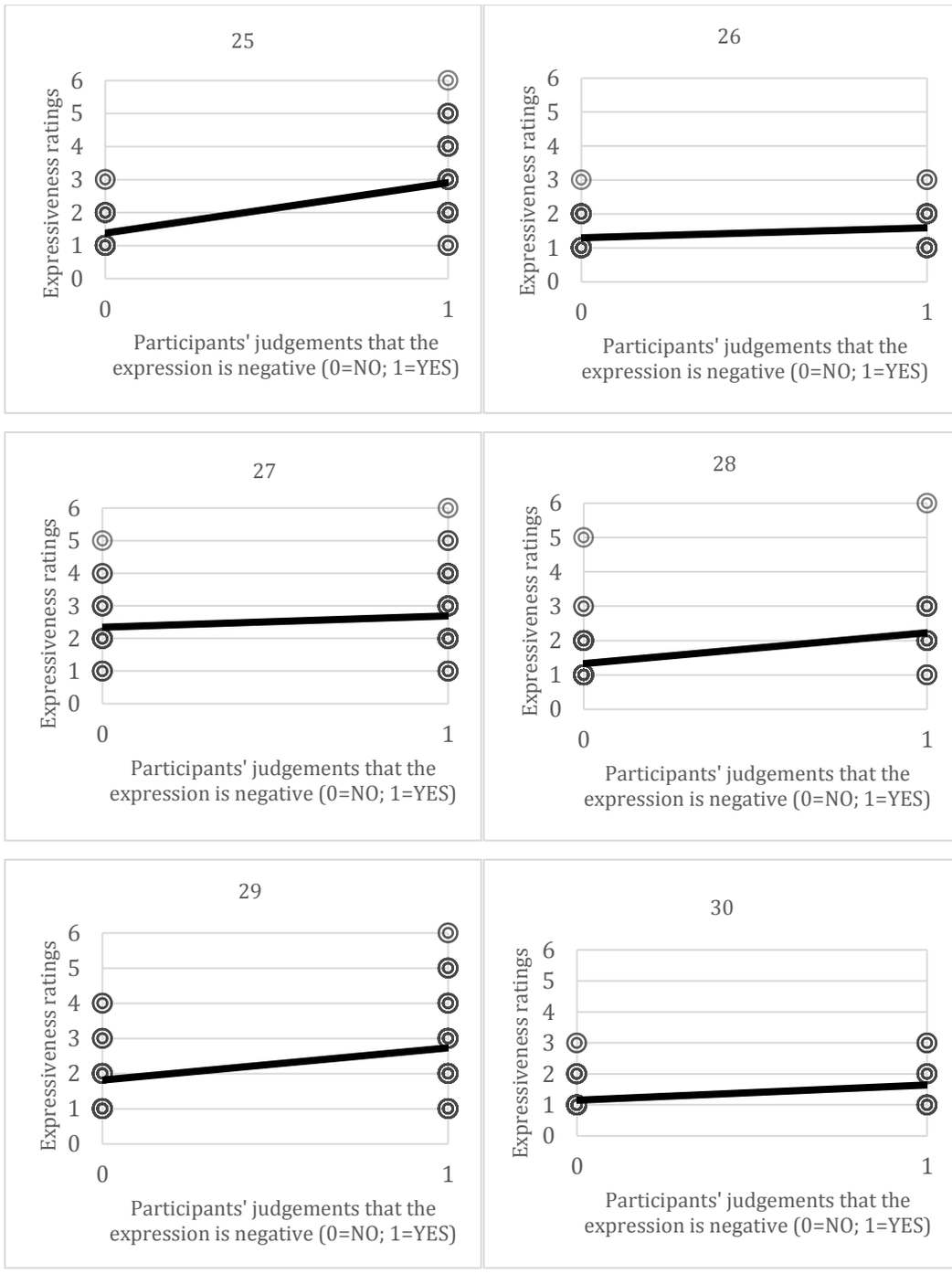

15
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)


16
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

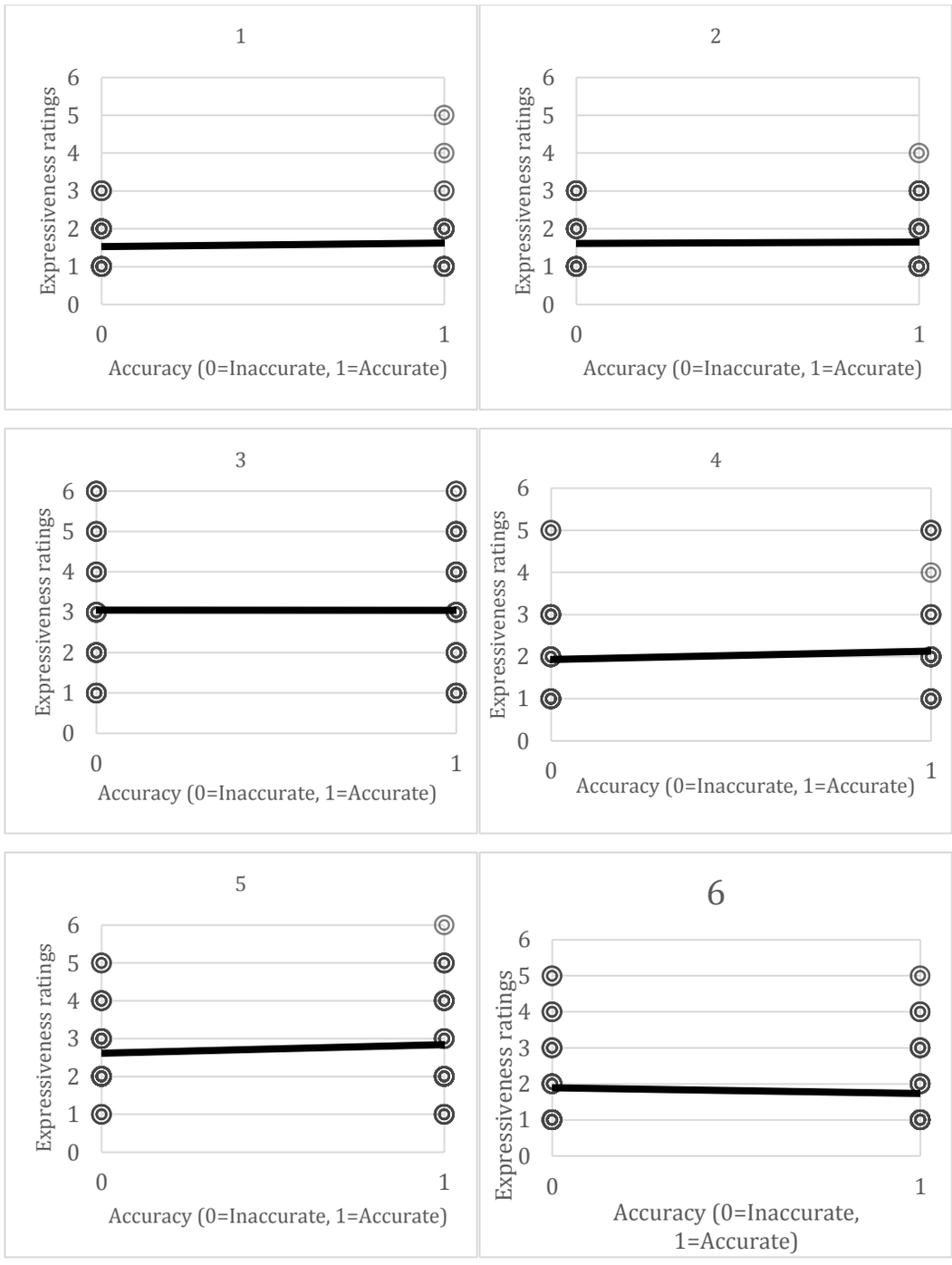

17
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

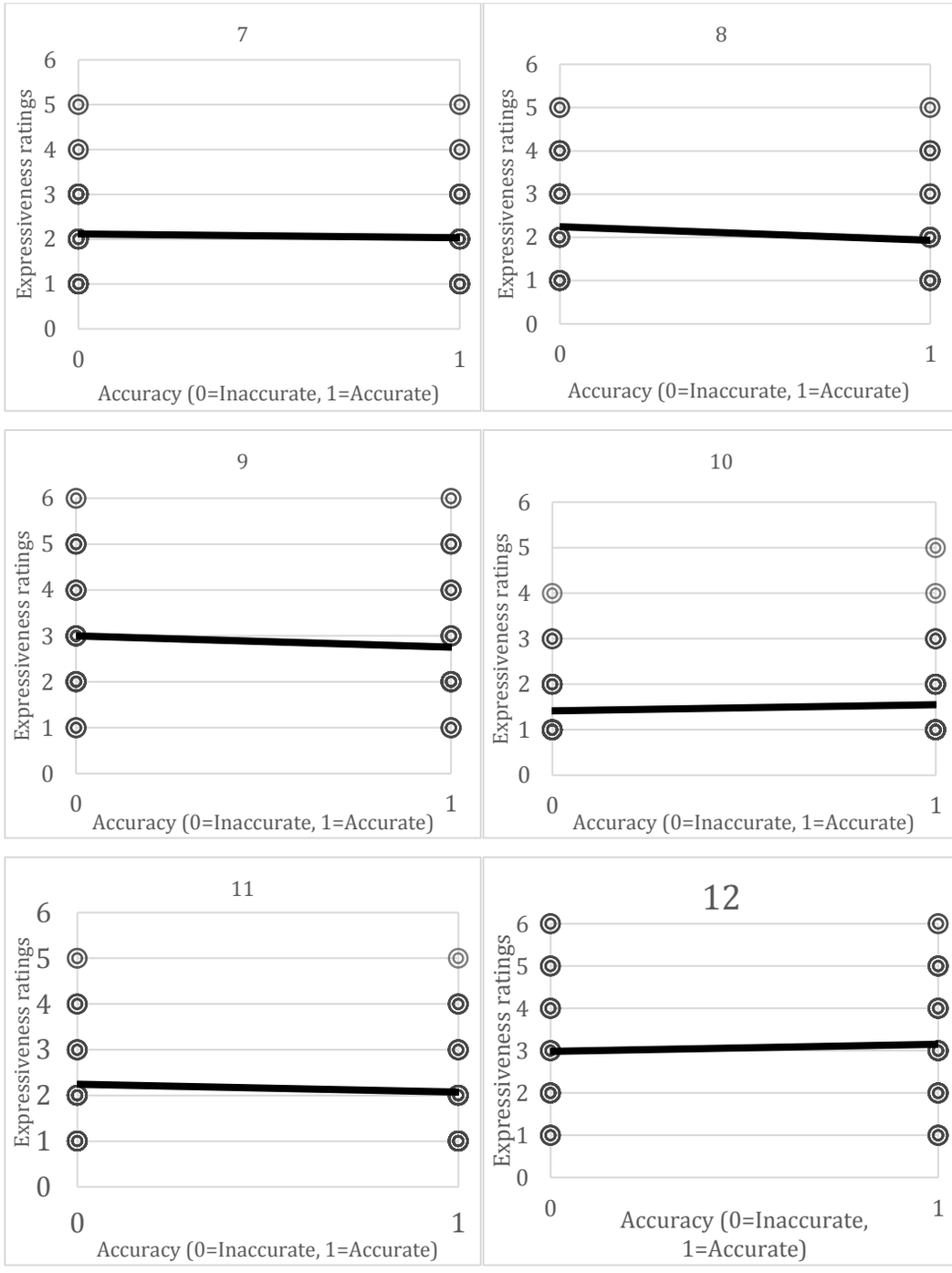

18
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

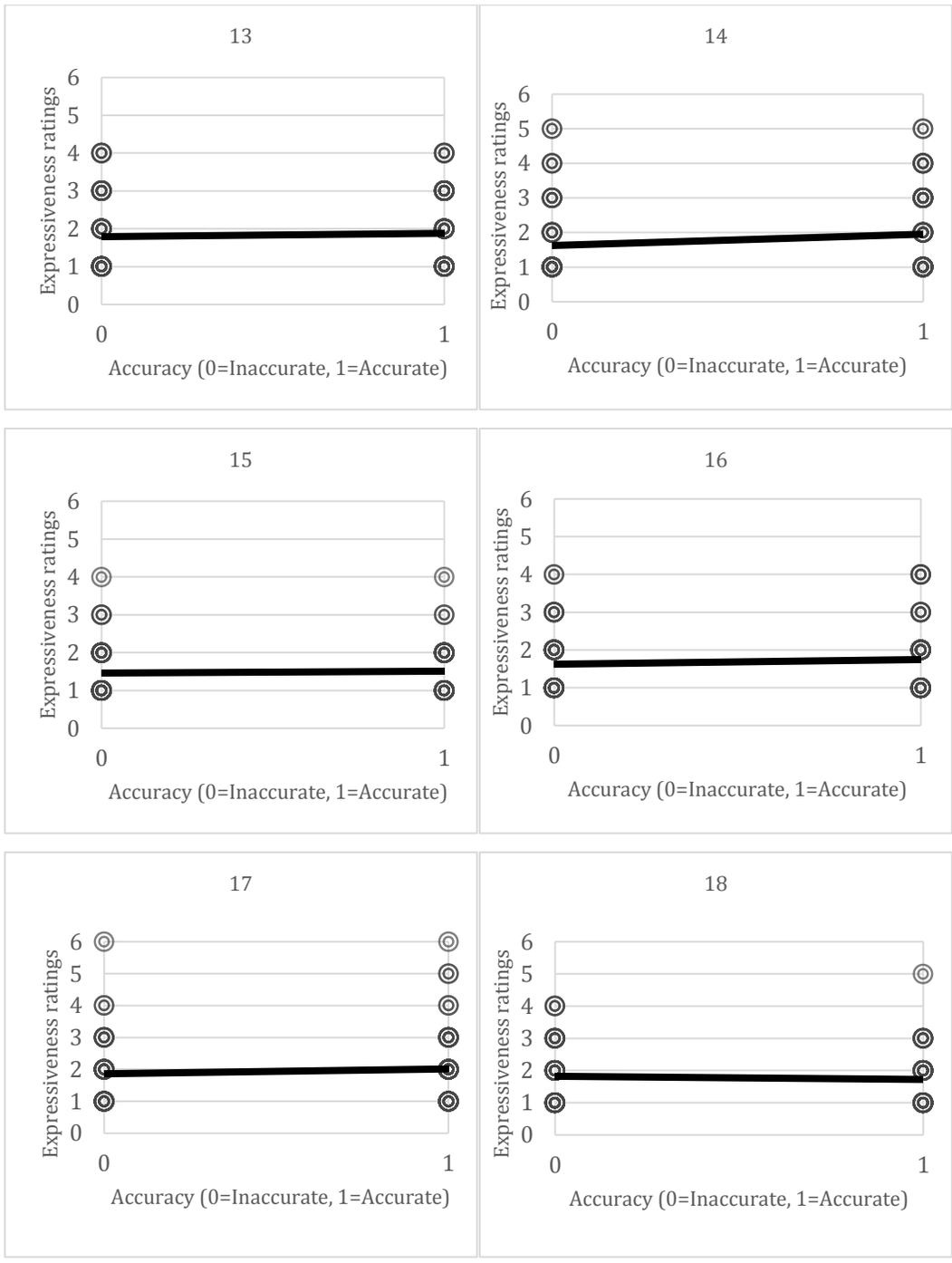

19
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)
20
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)
21
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)
22
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)
23
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)
24
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Participants' judgements that the expression is negative (0=NO; 1=YES)

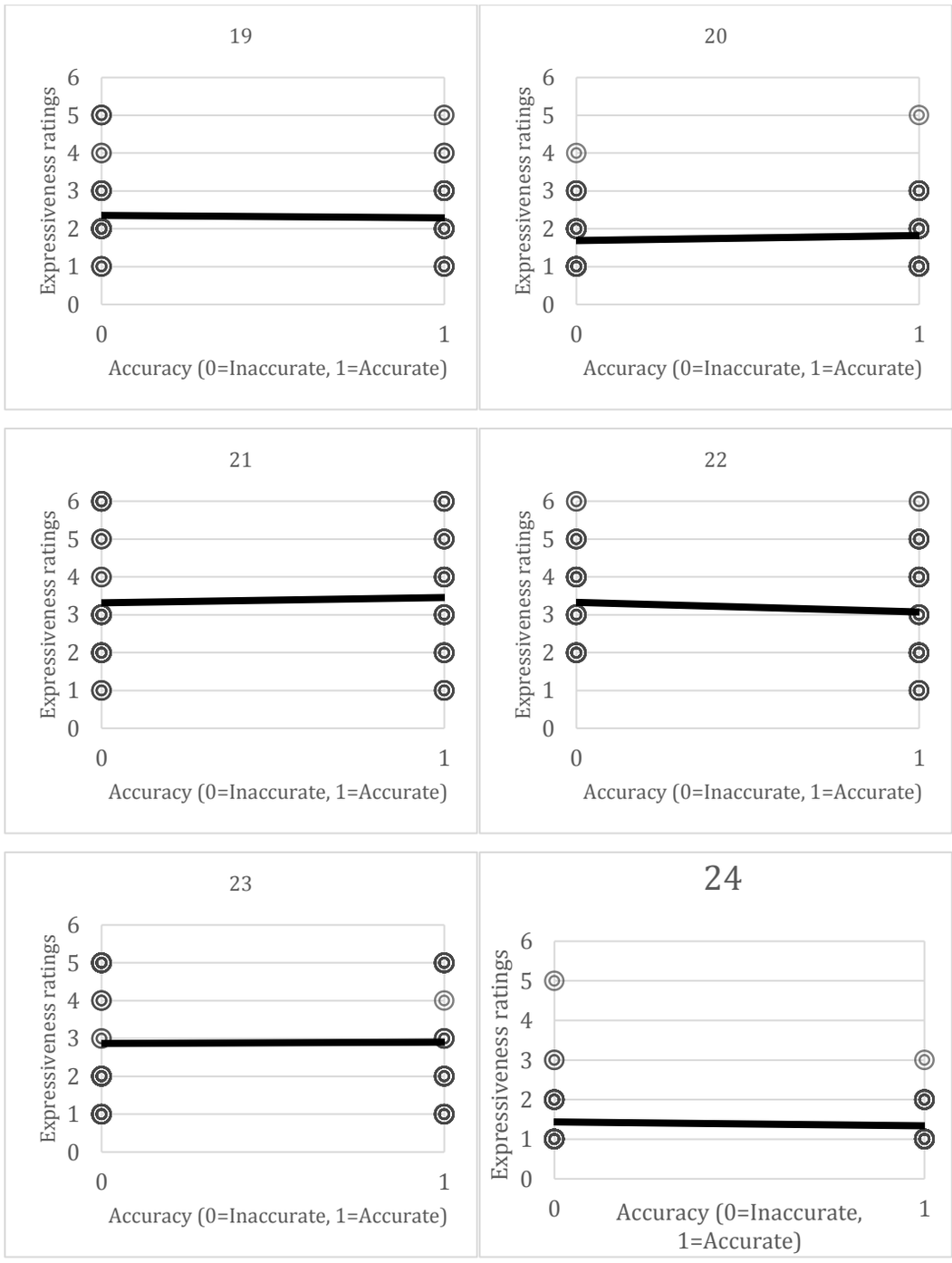


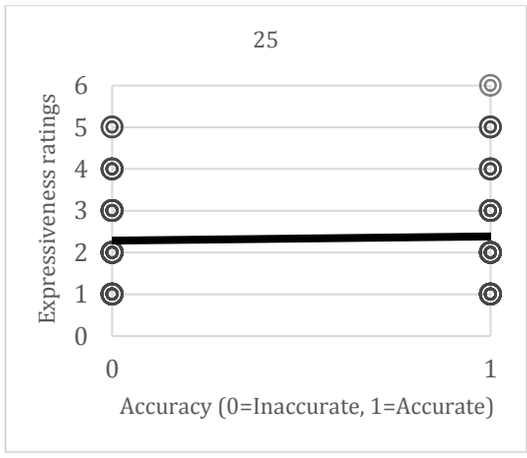
B. The relationship between expressiveness ratings and judges' accuracy, for all judges.

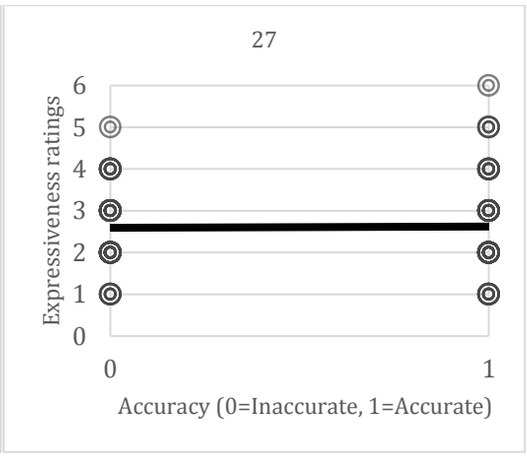

1
Expressiveness ratings
0
1
2
3
4
5
6
Accuracy (0=Inaccurate, 1=Accurate)
2
3
4
5
6
Accuracy (0=Inaccurate, 1=Accurate)

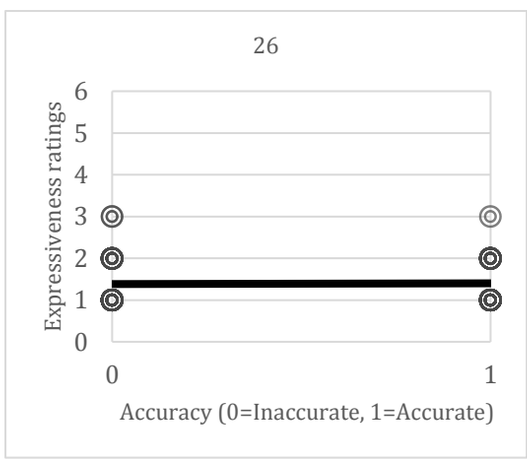

7
Expressiveness ratings
Accuracy (0=Inaccurate, 1=Accurate)
8
Expressiveness ratings
Accuracy (0=Inaccurate, 1=Accurate)
9
Expressiveness ratings
Accuracy (0=Inaccurate, 1=Accurate)
10
Expressiveness ratings
Accuracy (0=Inaccurate, 1=Accurate)
11
Expressiveness ratings
Accuracy (0=Inaccurate, 1=Accurate)
12
Expressiveness ratings
Accuracy (0=Inaccurate, 1=Accurate)

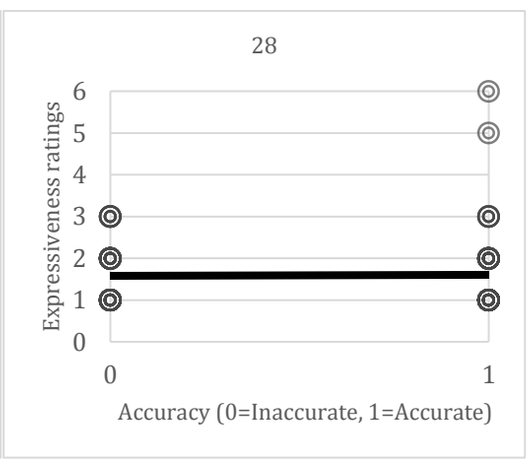

13
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Accuracy (0=Inaccurate, 1=Accurate)
14
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Accuracy (0=Inaccurate, 1=Accurate)
15
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Accuracy (0=Inaccurate, 1=Accurate)
16
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Accuracy (0=Inaccurate, 1=Accurate)
17
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Accuracy (0=Inaccurate, 1=Accurate)
18
Expressiveness ratings
6
5
4
3
2
1
0
0
1
Accuracy (0=Inaccurate, 1=Accurate)

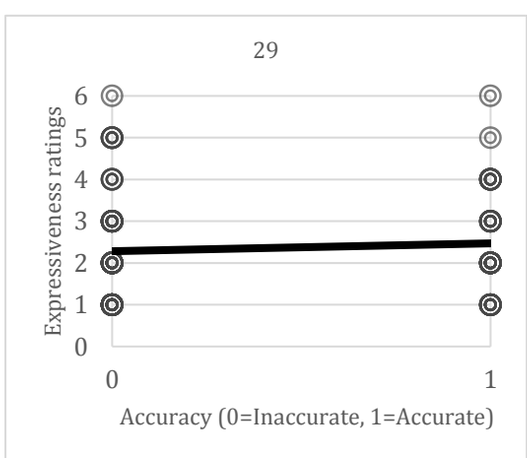

19
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)
20
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)
21
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)
22
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)
23
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)
24
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)

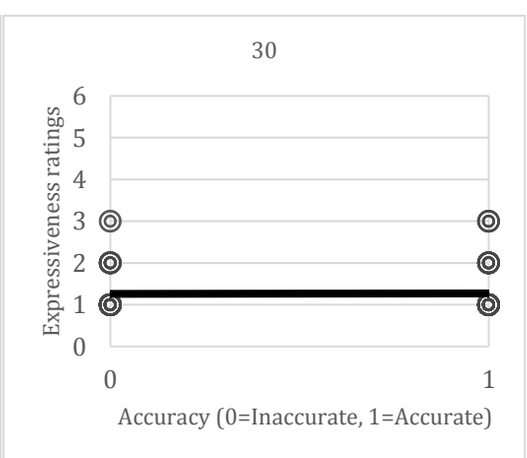

25
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)


27
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)


26
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)


28
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)


29
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)


30
Expressiveness ratings
0
1
2
3
4
5
6
0
1
Accuracy (0=Inaccurate, 1=Accurate)