Facial Disfigurement, Categorical Perception, and the influence of Disgust Sensitivity

Dr Anna Stone

University of East London

Author Note

Please contact Dr A Stone, School of Psychology, University of East London, Stratford, London E15 4LZ. Email: A.Stone@uel.ac.uk

Abstract

Previous research supports the categorical perception of faces on dimensions including emotion, identity, and gender. Two experiments using standard paradigms investigated whether facial disfigurement forms another perceptual category. In the Identification task, faces were presented in varying degrees of disfigurement for a simple disfigured / non-disfigured decision. As degree of disfigurement increased, the percentage of participants defining each image as disfigured increased non-linearly such that a cubic curve provided the best fit to the data, consistent with categorical perception (Experiment 1 and 2). In the ABX task, participants displayed superior discrimination between two faces when they crossed the category boundary between non-disfigured and disfigured (Experiment 1 and participants low in Disgust Sensitivity in Experiment 2). Participants high in Disgust Sensitivity (Experiment 2) showed a pattern that suggested the category boundary was shifted towards earlier perception of disfigurement. Overall, the results suggest categorical perception of facial disfigurement. *(148 words)*

Keywords: face; disfigurement; categorical perception; Disgust Sensitivity

Highlights

- Categorical perception of disfigured faces was investigated using standard paradigms
- Stimuli showed faces with position of one eye displaced or a large scarring on the cheek
- Results suggested that facial disfigurement is a perceptual category like identity or emotion
- Participants high in Disgust Sensitivity had earlier categorisation of facial disfigurement

Introduction

The charitable and campaigning organisation Changing Faces (2017) have estimated that in the UK, 540,000 people have a visible facial disfigurement impacting on others' perceptions and social behaviours. The automatic tendency to categorise individuals on easily detectable visual characteristics suggests it is important to know whether facial disfigurement is one such characteristic. There is a lack of research into the categorical perception of facial disfigurement and the present study aimed to provide some initial findings on this question.

A substantial body of research confirms that people with facial disfigurement suffer from discrimination. People with facial disfigurement are avoided (e.g., Jowett & Ryan, 1985; Lawrence, Rosenberg & Feuerbach, 2007; Rumsey, Bull & Gahagan, 1982; Ryan, Oaten, Stevenson, & Case, 2012) and invoke negative emotions (Bradbury, 2012; Rankin & Borah, 2003; Ryan et al, 2012; Shanmugarajah, Gaind, Clarke, & Butler, 2012; Stone & Colella, 1996; Stone & Potton, 2014). Generally, anomalous faces have been shown to give rise to negative and distrustful responses (Griffin & Langlois, 2006; Rankin & Borah, 2003). Other evidence shows that people with facial disfigurement are evaluated unfavourably on a variety of traits related to social competence and emotional strength (e.g., Bell & Klein, 2001; Bull & David, 1986; Hebl, Tickle, & Heatherton, 2000; Rankin & Borah, 2003; Stone & Wright, 2012) and may experience discrimination in an employment setting (e.g., Madera & Hebl, 2012; Stevenage & McKay, 1999; Stone & Wright, 2014). There is also evidence that facial disfigurement may be assumed to imply the possible presence of disease (e.g., Ryan et al, 2012; Shanmugarajah et al, 2012). Clearly, the presence of a visible facial disfigurement could have a major impact on the opportunities available to an individual.

Categorical perception occurs when individuals or objects are placed into pre-defined conceptual categories based on easily detectable visual characteristics. Categorical perception is often spontaneous and unconscious, arising without conscious deliberation (e.g., Bennett, Sani, Hopkins, Agostini, & Malucchi, 2000; Crisp & Hewstone, 2007; Habibi & Khurana, 2006) though limits to automaticity have been proposed (e.g., Freeman, Ma, Barth, Young, Han & Ambady, 2015).

The significance of categorical perception is that when the perceiver assigns the target to a category, this allows access to a set of stereotypical beliefs and expectations based on category membership. These beliefs and expectations can guide subsequent perceptions and behaviours towards the individual (e.g., Zebrowitz & Montepare, 2008) thus serving the purpose of cognitive efficiency. In the case of facial disfigurement, categorisation may lead to poorer evaluations, negative attitudes, and weaker intention to interact with the stigmatised individual. By analogy with racial discrimination, when a person is categorised as a member of a devalued social group, a set of automatic associations is activated with negative impact (e.g., Djik, 1989; Entman, 1992).

There is a substantial body of evidence supporting the proposition that human faces are categorised on a number of dimensions. Etcoff and Magee's (1992) influential study on categorical perception used line drawings of facial expressions of emotion as stimuli. The end points of the continuum were expressions of pure happiness or sadness and the faces showed a linear progression between these end points. Their finding was that participants switched abruptly between the perception of happiness and sadness, an effect taken to be a hallmark of categorical perception. Since then, experiments have revealed that categorical perception of faces occurs for other emotional expressions (e.g., Calder, Young, Perrett, Etcoff, & Rowland, 1996; Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002; de Gelder, Teunisse & Benson, 1997; Young, Rowland, Calder, Etcoff, Seth, & Perrett, 1997), and for the invariable aspects of identity (Angeli, Davidoff, & Valentine 2008; Beale & Keil, 1995; Levin & Beale, 2000; Viviani, Binda, & Borsato, 2007), ethnicity (Levin & Beale, 2000; Zhao & Bentin, 2008), and gender (Campanella, Chrysochoos, & Bruyer, 2001; Zhao & Bentin, 2008).

Given this evidence of facial categorisation on other dimensions, it seems plausible that individuals could be categorised on the easily detectable attribute of facial disfigurement. The impact of facial attractiveness on impression formation (e.g., Eagly, Ashmore, Makhijani, & Longo, 1991) supports the potential for categorisation based on facial appearance. No research that directly examined this question could be located but the study by Stone and Potton (2014) showed that under some circumstances, the emotional reactions to faces differed qualitatively between unattractive faces and disfigured faces. This suggests that disfigured faces may form a distinct perceptual category.

Several researchers have theorised that the human face processing system computes the average, or prototypical, face from a population of observed faces (e.g., Langlois & Roggman, 1990; Valentine, 1991; Valentine, Lewis & Hills, 2014). Valentine (1991) described a theoretical model of face space in which faces are encoded as points in a multi-dimensional space, where each dimension can be used to discriminate faces. In this model, the average or prototypical face is located in the centre of face space and unusual or atypical faces are located further from the centre. A disfigured face would be unusual and therefore would be located some distance from the centre of face space.

Several lines of research attest to the ability of the human face processing system to compute averages from the population of observed faces. For example, differential brain activation to disfigured faces compared to more typical faces establishes that the distinction between these faces is recognised in the visual system (e.g., Hartung, Jamrozik, et al, 2019; Parsons, Young, Mohseni, et al, 2013). The dependence of attractiveness evaluations on facial averageness (e.g., Langlois & Roggman, 1990; Potter & Corneille, 2008; Rhodes, Halberstadt, Jeffery, & Palermo, 2005) also implies that facial averages are detected in the visual system.

Facial prototypicality or averageness is associated with greater processing fluency (e.g., Trujillo, Jankowitsch, & Langlois, 2014; Principe & Langlois, 2012) and this is believed to underlie the preference for average faces. Similar effects are found in other stimuli more generally, e.g. Winkielman, Halberstadt, Fazendeiro, and Catty (2006) presented evidence that random-dot patterns and geometric patterns were processed more fluently (categorised more quickly) when they were closer to a prototype, and similar results were reported by Constable, Bayliss, Tipper, and Kritikos (2013).

In contrast with prototypical or average faces, disfigured faces are less familiar and further from the facial prototype. Hence, they will be processed less fluently by the visual system. The lack of familiarity and lack of fluency may be what underlies the categorisation of faces as disfigured.

The two experiments described here each used two standard paradigms for investigating categorical perception. The prediction was that both paradigms, implemented in the

Identification task (Task 1) and the ABX task (Task 2), would yield a pattern of responses indicative of categorical perception.

Experiment 2 also investigated the potential influence of Disgust Sensitivity on the categorisation of disfigured faces. Participants higher in Disgust Sensitivity would be predicted to be more sensitive to the presence of facial disfigurement and more vigilant in detecting facial disfigurement, and hence would define disfigurement at an earlier point of the continuum. This prediction follows from the function of disgust as a protective mechanism whose job is to detect and react to stimuli posing a threat of potential contamination (e.g., Rozin, Haidt & McCauley, 1993; Tybur & Leiberman, 2016). The disgust system tends to be over-inclusive (e.g., Park, Faulkner, & Schaller, 2003) and responds to facial differences posing no threat of contamination, hence the disgust system could be invoked by the perception of a facial disfigurement. Individuals higher in Disgust Sensitivity show more emotional reactivity, including disgust, to the perception of disfigured faces (e.g., Stone & Potton, 2014; Stone & Potton, 2017). This heightened reactivity would result in aversive reactions at lower levels of facial disfigurement and hence to an earlier category boundary for categorising a face as disfigured.

Experiment 1

Identification task

This used the standard paradigm in which two versions of the same face are blended in varying proportions to form a continuum. In the present study, the end-points of the continuum were non-disfigured and disfigured versions of the same face. Eleven versions of each face were created, ranging from 0% disfigured to 100% disfigured in 10% increments. Categorical perception would be indicated if participants reported that the stimulus changed abruptly from non-disfigured to disfigured, usually around the mid-point of the spectrum, in a cubic (s-shaped) curve, rather than changing gradually in smooth linear progression.

Method

Participants

There were 113 participants, of whom 82 were female, 27 male, and 4 declined to specify. Their ages ranged from 18 to 62 with a mean of 27.1 years and SD=8.9. The participants were recruited via various social media sites such as Facebook, Tumblr, and Instagram. The only specific requirement to participant was to be over the age of 18 years. None of the participants were paid for their participation. The participants came from a variety of ethnic backgrounds: 46 were Caucasian (British, Irish, other Caucasian), 47 were Black (Caribbean, African, other Black), 4 were Asian (Chinese, Japanese, Indian, Pakistani, Bangladeshi, other Asian) and 9 Other (Hispanic, Latino, Mixed, other). The remainder of the participants declined to specify their ethnicity.

A further three participants were excluded whose responses suggested a random pattern or the excessive use of a single response option.

According to Goode, Ellis, Coutinho, and Partridge, (2008) in a large study of 1000 adults, discrimination against people with facial disfigurement was not related to gender, socioeconomic status, or age.

Design

There was a single within-participant factor for task 1: degree of facial disfigurement. Eleven versions of each face were created by blending the non-disfigured version with the disfigured version in 10% increments from 0% to 100% disfigured. There were four face identities: male with eye disfigurement, female with eye disfigurement, male with mouth disfigurement, and female with mouth disfigurement. The dependent variable was the nominal response to a simple question: "is this face disfigured: Y or N". Each participant viewed every version of every facial identity three times, for a total of 132 trials.

Materials

Each stimulus was created using GIMP 2.8 by blending, in defined proportions, a non-disfigured face image and a disfigured face image of the same person. The images were those previously used in Stone and Wright (2012). To enable the blend to produce natural-looking images, 150 points were manually positioned on the face stimuli focusing on the significant points of the face and outlining all the internal features. The images used in Experiment 1 were taken from real individuals morphed with other faces to disguise their identity. Hence, they represent real disfigurements, and though they may be considered somewhat extreme, they are plausible. The faces differed in their gender, the nature of the disfigurements, the extremity of the disfigurements, and the attractiveness of the non-disfigured individuals. Hence, the results are shown for the four facial identities averaged together.

Figure 1 about here

Procedure

A pilot study verified that the stimuli would not cause discomfort to participants and tested the instructions and the expected duration of the procedure. The procedure was presented online and participants completed at times and locations of their choosing. The faces were presented one at a time in a single sequence that randomised all 44 faces, with the constraint that no face version was shown twice in succession. Participants were asked to decide whether each face was disfigured or not and to respond by clicking on “Yes” or “No”. When the response was received, the next face was presented after an inter-trial interval of 1 second. A practice set of 6 faces was included prior to the experimental trials so that the participants could view the stimuli, determine if they were comfortable participating, and become familiar with the procedure.

Ethical Considerations. This study was ethically approved by the ethics committee of University of East London. Before starting the procedure, participants were informed that the purpose of the study was to look at perceptions of facial disfigurements. They were assured that the results would be anonymous and that they could cease participation at any time and could withdraw their data up to one week from the date of participation. At the end of the procedure, participants were given information about the experiment and the email address of the experimenter with an invitation to make contact if they had any further questions about the experiment.

Results

Each participant viewed each of the 44 face stimuli (11 versions of 4 facial identities) 3 times. The three trials of the same stimulus were averaged for each participant and the result ranged from 0 to 1 (as the only possible responses were 0 or 1). There were no missing values. Participants whose responses suggested a random pattern (i.e., the likelihood of defining a face as disfigured showed no progression with increasing degree of disfigurement) or the use of a single response key, were excluded from the analysis.

The a-priori expectation was that each participant would start with “no” responses, and at some point around the approximate mid-point of the continuum would switch to “yes” responses, allowing for some variability around the crossover point. If disfigured faces are perceived categorically then the cumulative "yes" responses should be in the form of an S-curve with two points of inflexion, as the number of "yes" responses is initially very low, then rises steeply around the centre of the continuum, and finally levels off. Alternatively, if disfigured faces are not perceived categorically, then the cumulative "yes" responses should show a linear progression with increasing level of disfigurement.

Inspection of Figure 2, panel A suggests an S-curve (cubic curve) typical of categorical perception with the point at which 50% of participants identified a face as disfigured lying at approximately the 36% disfigurement point on the x-axis for the curve showing all four faces combined. The S-curve for each of the facial identities are shown in Figure 2 and these lie close together, suggesting a similar category boundary. The presence of categorical perception was investigated by examining the goodness-of-fit of both a linear function and a cubic function to the mean cumulative "yes" responses. The cubic function gave a better fit to the data than a linear function: $F(3,7) = 81.05$, $p<0.001$, R-square = 0.97 for the cubic function compared to $F(1,9) = 65.94$, $p<0.001$, R-square = 0.88 for the linear function. The addition of the third-order component improved the degree of fit, $\beta= 2.28$, $t(10)=2.64$, $p<0.05$. This pattern of results supports the presence of categorical perception of faces with disfigurement.

Figure 2 about here

Discussion

The results of the Identification task were consistent with categorical perception of the faces, showing an abrupt transition between identification of a face as non-disfigured and identification as disfigured as the degree of disfigurement increased linearly. A cubic curve showed the best fit to the data; this is a hallmark of categorical perception.

The crossover point between the non-disfigured and disfigured categories occurred at around the 36% disfigurement point for the four faces combined. The crossover point was not at the 50% blend, as is often the case, but this is not without precedent. For example, Calder et al (1996) found crossover points for categorical perception of sad vs. angry faces, and angry vs. fearful faces, that were not at 50% blend. It should also be noted that there is no defined end-point for a disfigured face, as there would be, for example, in a continuum that ranged between two facial identities, or facial expression, or ethnicities. Hence, there is no prior assumption that the cross-over point would be at the 50% blend.

ABX task: Method

In the standard paradigm a pair of faces is presented sequentially, and then one of the faces is presented again for a forced-choice decision; was this the first or the second face. Categorical perception is demonstrated if responses are faster and/or more accurate when the faces in the pair cross the category boundary than when both faces come from the same category. Applied to the present study, it would be easier to discriminate between a face from the non-disfigured side of the category boundary and a face from the disfigured side of the boundary than between two faces from the non-disfigured category or two faces from the disfigured category. Note that in each case the faces in the pair should differ by the same distance along the continuum.

Because the crossover point discovered in the Identification task lay at approximately 36% along the disfigurement continuum, the blended pairs for task 2 (ABX) were 0% and 18% (low-disfigurement), 27% and 45% (middle pair), and 54% and 72% (high-disfigurement). These numbers were chosen to present a consistent 18% difference between the two blends in each pair and to locate the crossover point at 36% in the centre of the middle pair. (Other studies in which the crossover point has been at around 50% have presented blended pairs of 10% and 30%, 40% and 60%, 70% and 90%. The pairs in the present study were adapted from the more usual pairs in such a way as to preserve the key features of the design).

Participants

There were 165 participants with valid data, comprising 109 female and 56 male, with ages ranging from 18 to 58 years, mean = 27 (S.D. = 9.0). Participants came from variety of ethnic backgrounds; 115 were Caucasian (British, Irish, other), 20 were Black (Caribbean, African, Other), 26 were Asian (Chinese, Japanese, Indian, Pakistani, Bangladeshi, other) and the remainder were Hispanic, Latino, or Mixed. A further 13 participants' data were excluded, five because their accuracy was below 45%, and eight because more than 20% of their response times were outliers.

Design

The independent variable was stimulus pair, with three levels of low-disfigurement 0% and 18%), middle pair (27% and 45%), and high-disfigurement (54% and 72%). The dependent variables were accuracy and response time in the ABX forced choice response.

On each trial, the two faces from a stimulus pair were presented sequentially and then one of these faces was presented again. The participant made a forced-choice decision whether the third face was identical to the first or the second face. For example, a trial might present the 45% blend, then the 27% blend, and finally the 45% blend again (BAB trial). Every participant performed 96 trials, comprising two instances of each combination of the three stimulus pairs x four face identities x four trial types (ABA, ABB, BAA, and BAB). Accuracy and response time were calculated for each stimulus pair by averaging over the other factors.

Materials

The new stimulus faces were prepared using the same techniques as the original faces, selecting the new blend levels of 18, 27, 45, 54, and 72 instead of the 10% intervals.

Procedure

Ethical considerations were similar to task 1.

After indicating consent, participants were presented with a practice phase consisting of four of the stimulus pairs. The instructions told them that they would see two faces and then a third face that was one of the original two. They were to hit “Z” if the third face was the same as the first face or “M” if the third face was the same as the second face. This allowed participants to familiarise themselves with the procedure, the duration for which each face was presented, and the nature of the stimuli, before proceeding with the experimental trials.

The instructions were repeated immediately before the first experimental trial. The order of the 96 trials was randomized. The timings for each trial were as follows: fixation cross for 500ms, blank screen for 250ms, the first image for 500ms, inter-stimulus interval of 250ms, second image for 500ms, inter-stimulus interval of 250ms, and final image for 750ms (based on Calder et al, 1996, with each stage decreased by 250ms to shorten the running time). The final test image was displayed for 750ms to encourage timely responding, on the consideration that responses were unlikely to become any more accurate if participants were given more time to study the final image. Evidence from studies of neural responses to disfigured faces (e.g., Huffmeijer, Barak-Levy, & Rinne, 2020; Parsons et al, 2013) suggests that disfigured faces are distinguished from typical faces starting at 140ms from stimulus onset. Hence the presentation times of 500ms for faces A and B, and 750ms for face X, appear sufficient for the detection of facial disfigurement. The entire procedure took around 10 minutes to complete.

Results

Responses were coded as correct or incorrect and the response time was recorded in milliseconds from the onset of the third face. The requirement to respond on every trial ensured that there were no missing responses. After inspection of the data, responses slower than 4 seconds (mean + 2.5 SDs) or faster than 300 milliseconds (improbably fast) were excluded as outliers. Participants who had more than 20% of their responses excluded were omitted from the analysis and the remaining exclusions amounted to less than 2% of the data. Participants with accuracy below 45% were also excluded as this comes close to a chance level of performance. It might be suggested that the exclusion cut-off point for accuracy should be different in the middle pair and the other pairs, given that performance was predicted to be more accurate in the middle pair, but this presupposes the effect of categorical perception that the experiment sets out to investigate.

Categorical perception predicts that responses to the middle pair (27-45% disfigured, crossing the category boundary) would be more accurate and faster than responses to the low-disfigurement (0-18%) and the high-disfigurement (54-72%) stimulus pairs. Analysis of Variance (ANOVA) was performed with a single within-participant factor of stimulus pair (low-disfigurement, middle, high-disfigurement). The dependent variables were accuracy and response time. Paired-samples t-tests (with one-tailed p-values) were used to follow up a significant Anova result with Bonferroni adjustment for multiple tests. Please refer to Figure 2.

With accuracy as the dependent variable there was a significant main effect of stimulus pair, $F(2,163) = 54.48$, $p<0.001$. Paired-sample t-tests revealed higher accuracy for the middle pair ($M=0.72$) than for the low-disfigurement pair ($M=0.65$), $t(164) = 0.08$, $p<0.001$, and higher accuracy for the middle pair ($M = 0.72$) than for the high-disfigurement pair ($M=0.63$), $t(164) = 9.57$, $p<0.001$.

The results for response time mirrored the results for accuracy. There was a significant main effect of stimulus pair, $F(2,163) = 9.74$, $p<0.001$. Paired-sample t-tests revealed faster responses for the middle pair ($M=1360$) than for the low-disfigurement pair ($M=1388$), $t(164) = 2.14$, $p<0.02$, and faster responses for the middle pair ($M=1360$) than for the high-disfigurement pair ($M=1418$), $t(164) = 4.42$, $p<0.001$.

Discussion

The results of the ABX task confirmed the prediction of faster and more accurate responses to the middle pair, in which the faces crossed the category boundary, compared to the low-disfigurement pair and the high-disfigurement pair. It appears that the task of distinguishing between two faces is facilitated when one is identified as disfigured and the other is not. This supports the hypothesis that disfigured faces are perceived categorically.

One limitation of Experiment 1 was that the four disfigured faces showed relatively extreme forms of disfigurement. It is possible that this could have facilitated the observation of categorical perception. To overcome this limitation, Experiment 2 used stimuli with milder forms of disfigurement. A similar pattern of results in Experiment 2 would support the hypothesis that categorical perception of disfigured faces is a general phenomenon and not restricted to faces showing extreme forms of disfigurement.

Another limitation of Experiment 1 is the use of the 0% disfigured face as one of the stimuli in the ABX task. This is not normally done but was necessitated by the location of the category boundary at 36% disfigured. The change to stimuli with a milder form of disfigurement was expected to shift the category boundary towards the 50% disfigurement and so to remove the need to use the 0% disfigured face.

Experiment 2

Experiment 2 used the same methodology as Experiment 1, and in addition examined the influence of Disgust Sensitivity (DS) on the categorisation of disfigured faces. The prediction was for an earlier category boundary between non-disfigured faces and disfigured faces in participants higher in Disgust Sensitivity. This could be apparent in the results of the Identification task showing the cross-over point between the categories of disfigured and non-disfigured faces at an earlier point for participants higher in Disgust Sensitivity. In the ABX task, the earlier cross-over point would mean that for some participants high in Disgust Sensitivity, the low-disfigurement pair might, in fact, be a cross-category pair; consequently, the middle pair might actually present a pair of faces both categorised as disfigured. So, this would predict a shift of accuracy and speed from the middle pair to the low disfigurement pair for participants higher in Disgust Sensitivity. The participants low in Disgust Sensitivity were predicted to show the standard effect of categorical perception, that is, superior performance in the middle pair. Disgust Sensitivity was measured via the Sensitivity subscale of the Disgust Propensity and Sensitivity Scale of van Overveld, Jong, and Peters (2009).

A second difference was the creation of new stimuli to investigate categorical perception over a more subtle degree of disfigurement, and thus to investigate the generality of the effect.

The images used in Experiment 1 showed strong disfigurements and this might have facilitated the detection of categorical perception; a more subtle range of stimuli would help to establish that the effect of categorical perception of facial disfigurement is a general effect.

The new stimuli were designed to have a milder degree of disfigurement, on the expectation that this would shift the cross-over point between disfigured and non-disfigured categories closer to the 50% blend. This would avoid the use of the 0% disfigured face as one of the stimuli in the ABX task.

Hypothesis 1 predicted that, in the Identification task, the crossover point between categorising faces as non-disfigured vs. disfigured might occur at a lower blend of disfigurement for those participants higher in Disgust Sensitivity compared to those low in Disgust Sensitivity. This would predict a negative correlation of Disgust Sensitivity with category boundary. Hypothesis 2 predicted that, in the ABX task, participants low in Disgust Sensitivity would show the usual effect of categorical perception, that is, faster and more accurate responses to the middle (cross-category) pairs compared to the low-disfigurement and high-disfigurement (same-category) pairs. Hypothesis 3 predicted that, in the ABX task, participants high in Disgust Sensitivity would show a shift of accuracy and speed from the middle pair towards the low-disfigurement pair. This would predict a positive correlation of Disgust Sensitivity with the mean accuracy in the low-disfigurement pair minus that in the middle pair, and a negative correlation of Disgust Sensitivity with the mean response time in the low-disfigurement pair minus the mean response time in the middle pair.

Identification task: Method

This used the same standard procedure as Experiment 1.

Participants

There were 32 participants, 25 female and 7 male, aged between 18 and 53 years with a mean of 23.7 years, SD = 9.6. Participants were mostly undergraduate students, and mainly residing in the United States of America. No participants' data were excluded.

Design

The design was the same as Experiment 1.

Materials

Two male and two female staff at the University of East London gave consent to have their photographs taken and manipulated for the use of the study. The four target models were of an approximately similar age (40-55) and of undistinguished appearance. The images were taken on an iPhone 6s against the same white wall at similar times of day controlling for the presence of natural sunlight and were uploaded as a PDF. Each face was cropped so only minimal clothing, the collar of a shirt, was visible. Each face was cropped to 53.34 by 44.77 cm with a resolution of 72 ppi. Two of the faces, one male and one female, were given eye disfigurements by moving the eye to match images of eye disfigurements found on Changing Faces website. The other two faces, one male and one female, were given lower face and mouth disfigurements by overlaying burn scars from burn images and manipulating the mouth to mimic the images.

Preliminary testing was carried out in order to arrive at a set of four faces with equivalent levels of perceived disfigurement. Each of the four disfigured images was blended with the corresponding original image in different proportions to create three levels of disfigurement: low, medium, and high. In the preliminary ratings of the stimuli, there were 22 participants, 20 female and 2 male, aged between 20 and 23 years old. They were all university students from

Boston Massachusetts. Participants were asked to rate the 12 faces, 3 levels of each of the 4 stimuli, on a scale from 1 = no disfigurement visible to 10 = maximum disfigurement. The ratings were averaged across participants for each of the 12 faces. A set of four faces, one for each identity, were selected with equivalent perceived disfigurement, approximately 6 on the scale of 1 to 10.

These four faces became the disfigured end points of their corresponding continua. Each disfigured face was blended with its original, non-disfigured version in different proportions. Four sets of facial stimuli were created, each containing 11 faces ranging from 0% disfigurement to 100% disfigurement in 10% increments. This created the 44 stimulus faces. The degree of disfigurement in these faces was lower than in the faces from Experiment 1, to represent a less surprising set of stimuli to participants. Examples of stimuli are shown in Figure 3.

Figure 3 about here

Participants completed the Disgust Propensity and Sensitivity Scale (DPSS; van Overveld et al, 2009) after the experimental trials. Each question asks participants to think about “how often it is true for you” with responses of never, rarely, sometimes, often, and always. Examples of questions on the Disgust Sensitivity scale are “When I feel disgusted, I worry that I might pass out” and “Disgusting things make my stomach turn”. The 6 items measuring Disgust Sensitivity (the unpleasantness of the experience of disgust) were used to calculate the score on Disgust Sensitivity for each participant.

Procedure

The procedure was the same as Experiment 1 with the addition of the DPSS scale after the experimental trials.

Results

A visual inspection detected one participant whose responses followed a random pattern and this participant's data were excluded from the analysis.

Inspection of Figure 4 suggests an S-curve (cubic curve) typical of categorical perception. Looking at the mean of the whole sample, the point at which 50% of participants identified a face as disfigured lay at approximately the mid-point on the x-axis at the 50% blend. It appears that the use of more subtly disfigured faces in Experiment 2, compared to Experiment 1, was successful in moving the overall category boundary to the 50% point.

The presence of categorical perception was investigated by fitting a linear and a cubic function to the cumulative "yes" responses. The cubic function gave a better fit to the data than a linear function: $F(3,7) = 252.7$, $p<0.001$, R-square = 0.99 for the cubic function compared to $F(1,9) = 137.0$, $p<0.001$, R-square = 0.94 for the linear function. The addition of the third-order component improved the degree of fit, $\beta = 3.11$, $t(10)=6.32$, $p<0.001$. This observation supports the prediction based on the logic of categorical perception that there would be an s-curve with two points of inflection.

Figure 4 about here

It was theorised that a participant higher in Disgust Sensitivity would start to categorise faces as disfigured at an earlier point on the continuum, i.e., at a lower level of disfigurement (hypothesis 1). Selection of the earliest blend at which a face was categorised as disfigured was not straightforward because there was a degree of variation in individual participants' responses, for example, a participant might not give a consistent answer to the same blend for the four different faces. The point at which faces started to be defined as disfigured was calculated for each participant as the lowest blend at which a non-zero proportion of the faces were defined as disfigured as long as the proportion of the faces defined as disfigured increased thereafter for subsequent blends. There was a negative correlation between the degree of Disgust Sensitivity and the point at which participants started to identify disfigurement, $r(30) = -0.43$, $p<0.05$. This supports the proposition that participants higher in Disgust Sensitivity would detect facial disfigurement at an earlier point on the continuum.

Participants were divided into two groups by median split on the Disgust Sensitivity scores. In the low Disgust Sensitivity group, M = 10.69, SD = 1.99, and in the high Disgust Sensitivity group M = 16.25, SD = 1.64. Figure 4 shows the cubic curve for the high Disgust Sensitivity group and the low Disgust Sensitivity group, separately, which illustrates the earlier category boundary for the participants higher in Disgust Sensitivity.

Examining the four facial identities separately revealed the category boundary to be in a similar position for both the low Disgust Sensitivity group (boundary at 46, 44, 46, and 43% disfigured for the female-eye, female-mouth, male-eye, and male-mouth, respectively) and the high Disgust Sensitivity group (boundary at 58, 56, 58, and 53%). All these category boundaries fall within the range of the middle pair (40% and 60% disfigured). Therefore, the subsequent analysis of the ABX task considered all four facial identities averaged together.

Discussion

The observation of a sharp transition between identification of faces as non-disfigured to identification as disfigured, as the degree of disfigurement increased linearly, was consistent with categorical perception and replicated the results of Experiment 1. A cubic curve with two points of inflection provided the best fit to the data, supporting the conclusion that disfigured faces are perceived categorically. The observation of a correlation between the category boundary and the scores on Disgust Sensitivity, such that the category boundary occurred at an earlier point for participants higher in Disgust Sensitivity, support Hypothesis 1 which predicted that facial disfigurement would be detected earlier in the presence of higher Disgust Sensitivity.

ABX task: Method

The crossover point in categorising a face as disfigured (vs non-disfigured) in the whole sample occurred at around the 50% blend. Therefore, the stimuli chosen for the ABX trials were the 10% and 30% blends, the 40% and 60% blends, and the 70% and 90% blends. These are the most commonly selected blends in previous research into the categorisation of face stimuli.

Participants

There were 87 participants, 59 female and 28 male, aged between 18 and 66 years *M*=31 years, *SD*=13.0. Of these, 61 participants were resident in the United States of America, mainly undergraduate students. The other 26 participants were an opportunity sample from the United Kingdom drawn from the general population. None were students of the University of East London. Data from a further seven participants were excluded; five participants had accuracy below 45%, and two participants had more than 20% of their response times classed as outliers.

Design

There were two independent variables: the within-participant variable was the stimulus pairs (low-disfigurement 10% and 30%, middle 40% and 60%, high-disfigurement 70% and 90%) and the between-participant variable was the level of Disgust Sensitivity, divided into low and high by median split. The dependent variables were accuracy and response time in the ABX forced choice response.

Every participant performed 48 trials, comprising the three stimulus pairs x four face identities x four trial types (faces ABA, ABB, BAA, and BAB). Accuracy and response time were averaged over the four trial types and the two faces with the same disfigured feature.

Materials

The faces in the ABX task were the 10% and 30% blends, the 40% and 60% blends, and the 70% and 90% blends, for each of the four face identities. Participants completed the DPSS after the experimental trials. Disgust Sensitivity was calculated for each participant as the sum of the 6 items on the Sensitivity subscale.

Procedure

The procedure was the same as Experiment 1, except that the number of trials was reduced to 48 in order to shorten the procedure and thus facilitate recruitment of participants. Each participant saw each combination of stimulus pair (10% and 30%, 40% and 60%, 70% and 90%) x face identity x trial type (ABA, ABB, BAA, BAB) only once.

Results

Responses were coded as correct or incorrect and the response time was recorded in milliseconds from the onset of the third face. The requirement to respond on every trial ensured that there were no missing responses. After inspection of the data, responses slower than 6 seconds (mean + 2.5 SDs) or faster than 300 milliseconds (improbably fast) were excluded as outliers; this amounted to less than 2% of the data. (The cut-off point of 6 seconds was higher than the cut-off of 4 seconds in Experiment 1, consistent with the more subtly disfigured faces in Experiment 2 that resulted in increased difficulty and the higher response times). Two participants had more than 20% of their responses excluded and these participant’s data were omitted from the analysis. Five other participants with low accuracy (less than 45%) were also excluded.

Participants were split into high and low Disgust Sensitivity groups by median split. In the low Disgust Sensitivity group M = 10.00, SD = 1.94, and in the high Disgust Sensitivity group M = 16.5, SD = 3.40. This compares with values of M = 10.63, SD = 3.57 (Overfeld et al, 2010), M = 12.83, SD = 4.44 (Fergus & Valentiner, 2009) and M = 13.59, SD = 5.44 (McKay, Yang, Elhai & Asmundson, 2020). The means of these studies all lie between the low Disgust Sensitivity group and the high Disgust Sensitivity group of the present sample.

The mean accuracy and response time for each of the three stimulus pairs (low-disfigurement, middle, and high-disfigurement) was calculated and summarised over the four facial identities. Analysis of Variance (ANOVA) was performed with one within-participant factor of stimulus pair and one between-participant factor of Disgust Sensitivity (high vs. low). The dependent variables were accuracy and response time. Paired-samples t-tests (with one-tailed p-values) were used to follow up a significant Anova result with Bonferroni adjustment for multiple tests. Please refer to Figure 4.

With accuracy as the dependent variable, there was a significant main effect of stimulus pair, $F(2,84) = 3.69$, $p<0.05$, and an interaction of stimulus pair with Disgust Sensitivity, $F(2,84) = 9.1$, $p<0.001$. Inspection of Figure 4 suggests that the pattern showing highest accuracy for the middle pair was specific to the participants with low Disgust Sensitivity. Paired-sample t-tests confirmed that for the participants with low Disgust Sensitivity: accuracy was higher for the middle pair (M=0.60) than for the low-disfigurement pair (M=0.54), $t(40)=2.42$, $p=0.01$, and near-significantly higher for the middle pair than for the high-disfigurement pair (M=0.54), t(40)=2.01, p=0.0255. This is the usual effect of perceptual categorisation. In contrast, for the participants with high Disgust Sensitivity: accuracy was higher for the low-disfigurement pair (M=0.65) than the middle pair (M=0.57), t(45)=3.36, p<0.005, and equivalent for the middle pair and the high-disfigurement pair (M=0.56), t(45)=0.78, ns.

Comparing participants high versus low in Disgust Sensitivity, in the low-disfigurement pair the high Disgust Sensitivity participants were more accurate [t(85) = 4.81, p<0.001] while in the middle pair the high Disgust Sensitivity participants were (non-significantly) less accurate, i.e., the peak accuracy was shifted towards the low-disfigurement pair for the high Disgust Sensitivity participants. The superior accuracy of the low-disfigurement pair minus the middle pair was correlated with Disgust Sensitivity [r(86) = 0.37, p<0.005]. This pattern supports the prediction that the category boundary would be shifted earlier for participants high in Disgust Sensitivity.

With response time as the dependent variable there was a significant main effect of stimulus pair, F(2,84) = 8.23, p=0.001, suggesting faster responses for the middle pair than the low-disfigurement or high-disfigurement pairs. There was no interaction with Disgust Sensitivity, F(2,84)=1.27, ns, and inspection of Figure 4 suggests that the effect of faster responses for the middle pair was present for participants high or low in Disgust Sensitivity. Paired-samples t-tests showed that responses were faster in the middle pair (M=1456ms) than in the low-disfigurement pair (M=1519ms), t(86) = 2.97, p<0.005, and faster in the middle pair than in the high-disfigurement pair (M=1535ms), t(86) = 3.22, p<0.005.

The correlation of Disgust Sensitivity with the difference in response time between the low-disfigurement pair and the middle pair was negative, [r(86) = -0.26, p<0.05] showing a shift of faster performance towards the low-disfigurement pair as the Disgust Sensitivity scores increase. This shift appears weaker than the shift of accuracy but shows a similar relationship with Disgust Sensitivity.

Discussion

This pattern of results supports the prediction in hypothesis 2 of faster and more accurate responses to middle pairs, in which the two faces come from different sides of the category boundary, than to low-disfigurement or high-disfigurement pairs, for participants low in Disgust Sensitivity. This is the usual effect of categorical perception. For these participants the results of Experiment 1 are replicated, suggesting that categorical perception of facial disfigurement was not dependent on the particular stimuli used in Experiment 1, but is a general phenomenon.

For participants high in Disgust Sensitivity, both accuracy and response times showed a shift of superior performance towards the low-disfigurement pairs, as predicted in hypothesis 3. This is consistent with the prediction that the category boundary would occur at an earlier stage of disfigurement for participants higher in Disgust Sensitivity. The apparent discrepancy between the accuracy and response time results, showing highest accuracy in the low-disfigurement pair, but fastest response time in the middle pair, requires some consideration. It is likely that the exact location of the category boundary would have varied among participants. Some of the participants would have had a category boundary in the range of the low-disfigurement pair while for others the category boundary would have been in the range of the middle pair. If the cross-category benefit in accuracy for the participants with category boundary in the low-disfigurement range was stronger than the benefit for those with category boundary in the middle range, this would explain the higher accuracy in the low-disfigurement pair. If the reverse were true for the response time data, and the cross-category benefit for the participants with category boundary in the middle range were stronger than the benefit for those with category boundary in the low-disfigurement range, this would explain the faster response times in the middle pair. This explanation is speculative and must await future research, but it is a plausible account of the pattern of results.

General Discussion

The Identification task in both experiments showed the hypothesised effect of categorical perception. There was an abrupt transition between the perception of faces as non-disfigured and their perception as disfigured as the degree of disfigurement increased in a linear continuum. A cubic curve provided the best fit to the observed data, as in previous studies in categorical perception. Experiment 2 also measured participants' score in Disgust Sensitivity, defined as the degree to which an individual experiences negative psychological and physical reactions to a disgust-inducing stimulus. As Disgust Sensitivity increased the category boundary occurred at a lower level of facial disfigurement, as predicted.

The ABX task showed the expected effects of categorical perception in Experiment 1 and for participants low in Disgust Sensitivity in Experiment 2. That is, a requirement to discriminate between two faces yielded faster and more accurate responses when the faces came from opposite sides of the category boundary between non-disfigured and disfigured (i.e., one non-disfigured and one disfigured face) than when they came from the same side of the boundary (either both non-disfigured or both disfigured).

For participants higher in Disgust Sensitivity (Experiment 2) a different pattern was observed. The correlation analysis demonstrated that the advantage in both higher accuracy, and faster response times, partially shifted from the middle pair towards the low-disfigurement pair as the scores on the Disgust Sensitivity questionnaire increased. This resulted in higher accuracy in the low-disfigurement pair than in the middle pair, though for response time, the overall advantage remained with the middle pair. This pattern of results is consistent with the explanation that the category boundary occurs at an earlier degree of disfigurement as the participant's level of Disgust Sensitivity increases, showing consistency with the results of the Identification task.

The four stimuli in Experiment 1 showed relatively extreme forms of disfigurement and this could have influenced participant responses and facilitated the observation of categorical perception. The four stimuli in Experiment 2 showed less severe forms of disfigurement and the same pattern of categorical perception was found (modified by level of Disgust Sensitivity). It appears that categorical perception of disfigured faces is a general phenomenon and not solely invoked by extreme forms of disfigurement.

This pattern of results can be interpreted in terms of the processing of in-group versus out-group faces. For example, Halberstadt, Sherman, and Sherman (2011) proposed that social categorisation is driven by the possession of traits of the minority group. We learn about common categories (typical faces) before we learn about uncommon or less frequent categories (e.g., disfigured faces) so that the process of learning how to detect members of the uncommon category can be done by attending to the features that distinguish a member of the uncommon category from the common category. So, we learn to categorise a face with a disfigurement by distinguishing it from a non-disfigured face. This increases the weighting given to the feature of the uncommon category when perceiving and categorising faces. The effect is to create a bias towards categorisation as a member of the uncommon category, so that when an ambiguous face is perceived, it is more likely to be assigned to the uncommon category.

It is proposed that this bias starts at a lower level of disfigurement for people higher in Disgust Sensitivity because these people would be more sensitive and reactive to the presence of disfigurement (e.g., van Overveld et al, 2009; Stone & Potton, 2014; Stone & Potton, 2017). This phenomenon, sometimes known as in-group over-exclusion, could be part of a motivation to protect the ingroup from contamination by members of a socially inferior group (e.g., Castano, Yzerbyt, Bourguignon, & Seron, 2002; Leyens & Yzerbyt, 1992).

Social categorisation of faces (e.g., by race, gender, or age) happens quickly and automatically (Hugenberg, Young, Bernstein, & Sacco, 2010). It follows that attention to the facial aspects that determine the category is rapid and automatic. There is also evidence that a disfigured feature captures attention (Stone & Potton, 2017; Madera & Hebl, 2012) and particularly so for people higher in Disgust Sensitivity (Stone & Potton, 2017). It seems likely that when viewing a disfigured face attention is paid mainly to the disfigured feature so the face is not processed in fine detail. This explanation would account for poorer accuracy for the task of distinguishing between two versions of a face with a perceived disfigurement. The poor discrimination between two disfigured versions of a face is similar in concept to the other race effect, in which two members of another race are more readily confused than two members of the perceiver's own race. A similar process may underlie both effects: the signifiers of race or disfigurement are processed, the face is categorised as other-race or disfigured, and is not processed in as much detail as an own-race or non-disfigured face.

Prototypical faces are perceived more frequently than distinctive faces and therefore may be processed more fluently. The differential processing fluency of disfigured faces compared to prototypical faces could be another mechanism underlying the categorisation of a face as disfigured. When the visual processing system detects a face with very different visual properties to the average face – a face that lies far from the centre of face-space (Valentine et al, 2014) - the face is flagged as unusual.

The overall pattern of results confirms the existence of categorical perception for facial disfigurement. This is consistent with numerous previous studies showing that faces are spontaneously categorised on easily perceived attributes of gender, emotion, ethnicity, and identity. Facial disfigurement is an easily perceived attribute, and given the importance of attractiveness to impression formation (e.g., Eagly et al, 1991) it is perhaps not surprising that faces should be categorised on the presence of disfigurement. It is also consistent with the observation from Stone and Potton (2014) that emotional reactions differed qualitatively between faces evaluated as unattractive and those evaluated as disfigured.

Since categorisation based on easily perceived characteristics can lead to the activation of stereotypical beliefs and expectations, this has the potential to explain negative reactions and behaviours towards individuals with facial disfigurement. If an individual is categorised as having a facial disfigurement this could result in the activation of assumptions about their personality traits and competencies. These assumptions are likely to be based on those depictions of people with facial disfigurement that are prevalent within the culture. It is also possible that expectations and beliefs might be based on actual acquaintance with someone who has a facial disfigurement but this seems rather less likely since previous studies have consistently found that only a minority of participants report having an acquaintance with facial disfigurement.

There are potential implications for equality of opportunity and non-discrimination in social and employment settings. The Equality Act (2010) offers protection to those who are severely disfigured, and the present research suggests a clear boundary between those who are regarded as disfigured and those who are not. A definition based on public perception, which is perhaps the key factor in the treatment an individual will receive, may be of benefit. It should be noted that the negative affect experienced by someone with a disfigurement is not always strongly related to the severity of disfigurement (e.g., Moss, 2005) but nonetheless a means of establishing who is or is not considered to be disfigured may have some utility.

Some limitations of this research should be noted. Only a small set of faces was employed, four in Experiment 1 and four more in Experiment 2, representing two types of disfigurement: asymmetrical location of the eyes and scarring on the mouth-cheek area. Although these are illustrative of common types of disfigurement, it would be useful to examine the impact of a wider range of disfigurement. Only static stimuli were employed and reactions to individuals displaying a range of movements might differ in some respects.

Disgust Sensitivity was self-reported after the experimental trials, so there is a possibility that experience of the faces could have influenced participants' responses on the Disgust Sensitivity scale. This would likely have the effect that the reported Disgust Sensitivity would have been rather specific to facial disfigurement and perhaps less generalised. The alternative, of recording Disgust Sensitivity before the experimental trials, was rejected because of the risk of priming participants with disgust concepts before seeing the faces. A future study might investigate whether the placing of the measure of Disgust Sensitivity has any influence on the pattern of categorisation.

Disgust Sensitivity was not recorded for the participants in Experiment 1, so analysis broken down by Disgust Sensitivity was not possible. The sample sizes in Experiment 1 (113 in the Identification task and 165 in the ABX task) suggest that participants were likely to have spanned a range of levels of Disgust Sensitivity and it is relevant to note the results of Experiment 1 were comparable to the overall results of Experiment 2.

Future research could use a wider range of stimuli and perhaps look for differences in the ways in which males and females are perceived. Although Stone and Wright (2012) suggested that commonly-observed gender differences in trait attributions are eliminated in faces with disfigurement, perhaps indicating that a disfigured face is less strongly gendered, there may still be some differential effects of gender on the categorisation of disfigurement. For example, the impact of disfigurement may be deemed to be more severe on a female face than on a male face, though this remains to be established.

Future research might also examine affective responses to individuals with facial disfigurement and how these relate to categorical perception. Affective responses could be the product of several factors including processing fluency, prototypicality/averageness, fear of disease, or assumptions about social skills and emotional stability. This lies beyond the scope of the present paper.

The conclusion can be drawn that faces with disfigurement are perceived categorically. For perceivers high in the trait of Disgust Sensitivity, there is a general tendency to categorisation of a face as disfigured at a lesser level of disfigurement.

The assistance of Kamile Stankeviciute, Donaleen Johnson, and Chisom Fadeyi (University of East London) and Madison Clement (University of San Francisco) in the data collection is gratefully acknowledged.

Conflicts of interest: None.

No financial support was received for the production of this work.

Data may be obtained by application to the author.

References

Angeli, A., Davidoff, J., & Valentine, T. (2008). Face familiarity, distinctiveness, and categorical perception. *The Quarterly Journal of Experimental Psychology*, 61(5), 690-707.

Beale, J.M., Keil, C.F. (1995). Categorical effects in the perception of faces. *Cognition*, 57, 217-239.

Bell, B. S., & Klein, K. J. (2001). Effects of disability, gender, and job level on ratings of job applicants. *Rehabilitation Psychology*, 46, 229–246.

Bennett, M, Sani, F., Hopkins, N., Agostini, L. & Malucchi, L. (2000) Children's gender categorisation: An investigation of automatic processing. *British Journal of Developmental Psychology*, 18, 97-102.

Blascovich, J., Mendes, W. B., Hunter, S. B., Lickel, B., & Kowai-Bell, N. (2001). Perceiver threat in social interactions with stigmatized others. *Journal of personality and social psychology*, 80(2), 253.

Bradbury, E. (2012). Meeting the psychological needs of patients with facial disfigurement. *British Journal of Oral and Maxillofacial Surgery*, 50(3), 193-196.

Bull, R., & David, I. (1986). The stigmatizing effect of facial disfigurement. *Journal of Cross-Cultural Psychology*, 17, 99–108.

Calder, A. J., Young, A. W., Perrett, D. I., Etcoff, N. L., & Rowland, D. (1996). Categorical perception of morphed facial expressions. *Visual Cognition*, *3*(2), 81-117.doi:10.1080/713756735

Campanella, S., Chrychoos, A. & Bruyer, R. (2001) Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition*, 8, 237-262.

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guerit, J. (2002). Categorical perception of happiness and fear facial expressions: An ERP study. *Journal of Cognitive Neuroscience*, *14*, 210-227.

Castano, E., Yzerbyt, V., Bourguignon, D., & Seron, E. (2002). Who may enter? The impact of in-group identification on in-group/ out-group categorization. *Journal of Experimental Social Psychology*, 38, 315–322.

Changing Faces (2017) *Disfigurement in the UK*. Retrieved from https://www.changingfaces.org.uk/campaigns/dituk on 16th August 2017.

Constable, M.D., Bayliss, A.P., Tipper, S.P., & Kritikos, A. (2013) Self-generated cognitive fluency as an alternative route to preference formation. *Consciousness and Cognition*, 22, 47-52.

de Gelder, B., Teunisse, J. P., & Benson, P. J. (1997). Categorical perception of facial expressions: Categories and their internal structure. *Cognition & Emotion*, 11(1), 1-23.

Eagly, A.H, Ashmore, R.D., Makhijani, M.G., & Longo, L.C. (1991) What is beautiful is good, but … A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, 110, 109-128.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227-240. doi:10.1016/0010-0277(92)90002-Y

Fergus, T.A. & Valentiner, D.P. (2009) The Disgust Propensity and Sensitivity Scale–Revised: An examination of a reduced-item version. *Journal of Anxiety Disorders*, 23 703-710.

Freeman, J.B., Ma, Y., Barth, M., Young, S.G., Han, S., & Ambady, N. (2015) The neural basis of contextual influences on face categorisation. *Cerebral Cortex*, 25, 415-422.

Goode, A., Ellis, R., Coutinho, W., & Partridge, J. (2008). The Face Equality Campaign – The Evidence: Public Attitudes Survey. London, England: Changing Faces. Retrieved from http://www. changingfaces.org.uk/downloads/FE Campaign, Public Attitudes survey.pdf

Griffin, A. M., & Langlois, J. H. (2006). Stereotype directionality and attractiveness stereotyping: Is beauty good or is ugly bad?. *Social Cognition*, 24(2), 187-206. doi:10.1521/soco.2006.24.2.187

Habibi, R. & Khurana, B. (2012) Spontaneous gender categorization in masking and priming studies: Key for distinguishing Jane from John Doe but not Madonna from Sinatra. *PLoS ONE*, 7, e32377, https://doi.org/10.1371/journal.pone.0032377

Hartung, F., Jamrozik, A., Rosen, M.E. et al. (2019) Behavioural and Neural Responses to Facial Disfigurement. *Scientific Reports,* 9, 8021. https://doi.org/10.1038/s41598-019-44408-8

Hebl, M. R., Tickle, J., & Heatherton, T. F. (2000). Awkward moments ininteractions between nonstigmatized and stigmatized individuals.In T. Heatherton, R. Kleck, M. Hebl, & J. Hull (Eds.), *The social psychology of stigma* (pp. 275–306). New York, NY: Guilford Press.

Huffmeijer, R., Barak-Levy, Y., & Rinne, R.C.A. (2020) Attractiveness and neural processingof infant faces: Effects of facial abnormality but not dopamine. *Physiology and Behaviour*, 222, doi:10.1016/jphysbeh.2020.112937

Hugenberg, K., Young, S.G., Bernstein, M.J., & Sacco, D.F. (2010) The Categorization-Individuation Model: An Integrative Account of the Other Race Recognition Deficit. *Psychological Review*, 117, 1168-1187.

Jowett, S. & Ryan, T. (1985) Skin disease and handicap: an analysis of the impact of skin conditions. *Social Science and Medicine,* 20, 425-429.

Langlois, J.H. & Roggman, L.A. (1990) Attractive Faces Are Only Average. *Psychological Science,* 1,115-121.

Lawrence JW, Rosenberg LE and Fauerbach JA (2007) Comparing the body esteem of pediatric survivors of burn injury with the body esteem of an age-matched comparison group without burns. *Rehabilitation Psychology* 52: 370–379.

Levin, D. T., & Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics*, 62(2), 386-401.

Leyens, J.-P., & Yzerbyt, Y. (1992). The ingroup overexclusion effect: Impact of valence and confirmation on stereotypical information search. *European Journal of Social Psychology*, 22, 549–569.

Madera, J. M., & Hebl, M. R. (2012). Discrimination against facially stigmatized applicants in interviews: An eye-tracking and face-to-face investigation. *Journal Of Applied Psychology*, *97*(2), 317-330. doi:10.1037/a0025799

McKay, D., Yang, H., Elhai, J. & Asmundson, G.J.G. (2020) Anxiety regarding contracting COVID-19 related to interoceptive anxiety sensations: The moderating role of disgust propensity and sensitivity. *Journal of Anxiety Disorders*, 73, https://doi.org/10.1016/j.janxdis.2020.102233.

Moss, T.P. (2005) The relationships between objective and subjective ratings of disfigurement severity and psychological adjustment. *Body Image,* 2,151-159.

Park, J.H., Faulkner, J. & Schaller, M. (2003) Evolved disease-avoidance processes and contemporary anti-social behaviour: prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior*, 27, 65-87.

Parsons, C.E., Young, K.S., Mohseni, H., Woolrich, M.W., Thomsen, K.R., Joensson, M., Murray, L., Goodacre, T., Stein, A., & Kringelbach, M.L. (2013) Minor structural abnormalities in the infant face disrupt neural processing: A unique window into early caregiving responses, *Social Neuroscience*, 8, 268-274.

Potter, T. & Corneille, O. (2008) Locating attractiveness in the face space: Faces are more attractive when closer to their group prototype. *Psychonomic Bulletin & Review* **15,** 615–622.

Principe, C. P. & Langlois, J. H. (2012). Shifting the prototype: experience with faces influences affective and attractiveness preferences. *Social cognition*, *30*, 109–120.

Rankin, M., & Borah, L. B. (2003). Perceived functional impact of abnormal facial appearance. *Plastic and Reconstructive Surgery*, 111, 2140-2146.

Rumsey, N., Bull, R., & Gahagan, D. (1982). The effect of facial disfigurement on the proxemic behavior of the general public. *Journal Of Applied Social Psychology*, *12*(2), 137-150. doi:10.1111/j.1559-1816.1982.tb00855.x

Ryan, S., Oaten, M., Stevenson, R. J., & Case, T. I. (2012). Facial disfigurement is treated like an infectious disease. *Evolution and Human Behavior*, 33(6), 639-646.

Shanmugarajah, K., Gaind, S., Clarke, A., & Butler, P. M. (2012). The role of disgust emotions in the observer response to facial disfigurement. *Body Image*, *9*(4), 455-461. doi:10.1016/j.bodyim.2012.05.003

Stevenage, S. V. & McKay, Y. (1999). Mo del applicants: The effect of facial appearance on recruitment decisions. *British Journal of Psychology, 90*, 221-234.

Stone, D. L., & Colella, A. (1996). A model of factors affecting the treatment of disabled individuals in organizations. *Academy of Management Review*, 21, 352–401.

Stone A and Potton A (2014). Emotional responses to disfigured faces: the influences of perceived anonymity, empathy, and disgust sensitivity. *Basic and Applied Social Psychology*, 36: 520-532.

Stone, A. & Wright, T. (2012). Evaluations of people depicted with facial disfigurement compared to those with mobility impairment. *Basic and Applied Social Psychology, 34*, 212-225.

Stone, A. & Wright, T. (2013). When your face doesn't fit: Employment discrimination against people with facial disfigurement. *Journal of Applied Social Psychology, 43*, 515-526.

Tybur, J.M. & Leiberman, D (2016) Human pathogen avoidance adaptations. *Current Opinion in Psychology*, 7, 6-11.

Valentine, T. (1991) A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology*: A, 43,161-204.

Valentine, T., Lewis, M. & Hills, P.(2014) Face-Space: A Unifying Concept in Face Recognition Research. *Quarterly journal of experimental psychology*, 69, 1-57.

van Overveld, M., de Jong, P. J., & Peters, M. L. (2010). The Disgust Propensity and Sensitivity Scale – Revised: Its predictive value for avoidance behavior. *Personality and Individual Differences, 49*, 706-711.

Viviani, P., Binda, P., & Borsato, T. (2007). Categorical perception of newly learned faces. *Visual Cognition*, *15*(4), 420-467. doi:10.1080/13506280600761134

Wegrzyn, M., Bruckhaus, I., & Kissler, J. (2015). Categorical perception of fear and anger expressions in whole, masked and composite faces. *Plos ONE*, *10*(8), e0134790. https://doi.org/10.1371/journal.pone.0134790

Winkielman, P., Halberstadt, J., Fazendeiro, T., & Catty, S. (2006) Prototypes are attractive because they are easy on the mind. *Psychological Science*, 17, 799-806.

Wong, A. C.-N., Palmeri, T. J., & Gauthier, I. (2009). Conditions for facelike expertise with objects: Becoming a Ziggerin expert – but which type? *Psychological Science*, 20, 1108-1117.

Young, A.W., Rowland, D., Calder, A.J., Etcoff, N.L., Seth, A., & Perrett, D.I. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271–313

Yovel, G., Halsband, K., & Goshen, Y. (2009). Dissociating between the role of exposure and individuation in perceptual expertise for faces. *Journal of Vision*, 9, 458.

Zebrowitz L.A., Montepare J.M. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. *Evolution and Social Psychology,* 81–113.

Zhao, L. & Bentin, S. (2008) Own- and other-race categorisation of faces by race, gender, and age. *Psychonmic Bulletin and Review*, 15, 1093-1099.

Figures

Figure 1: example of stimuli used in Experiment 1.

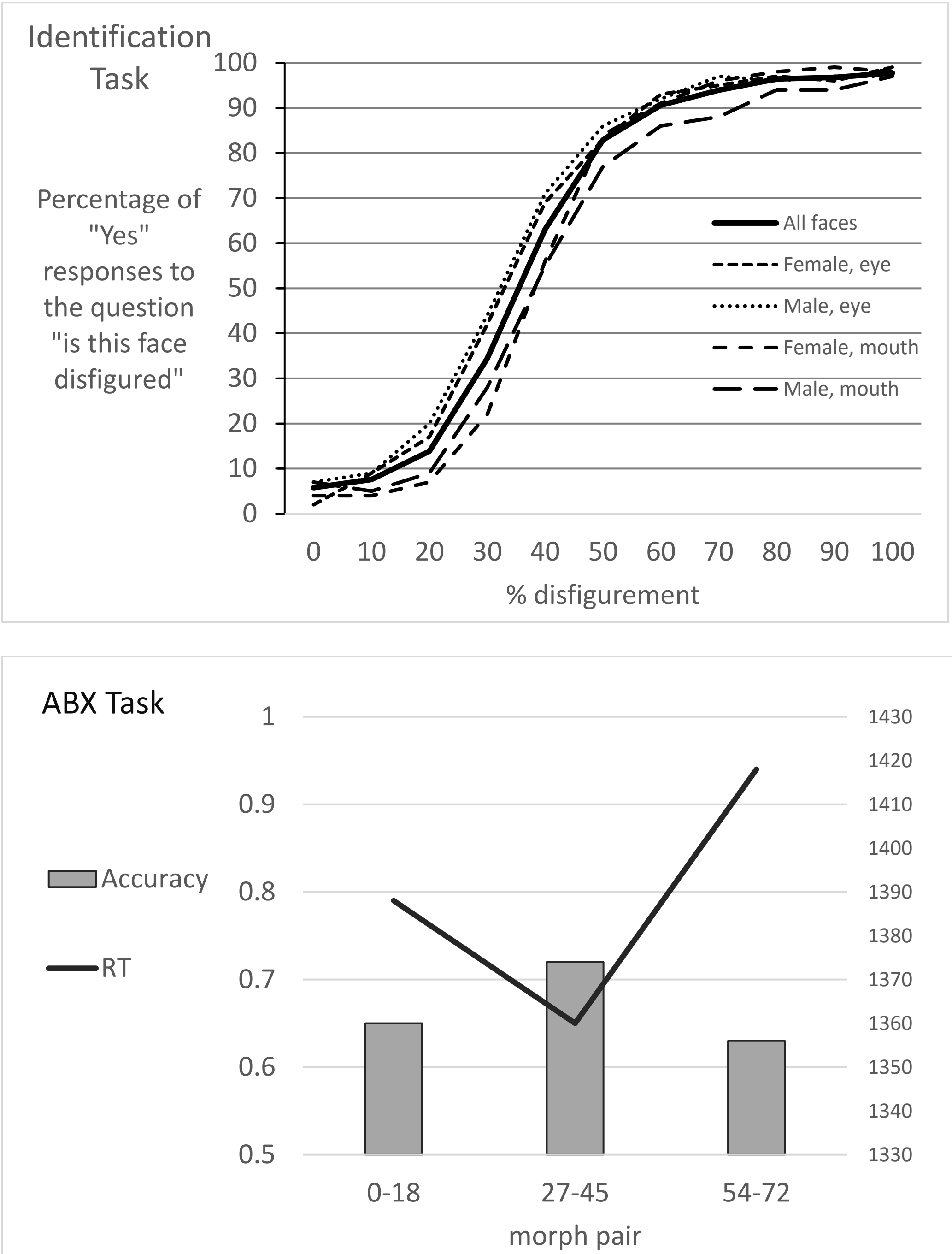
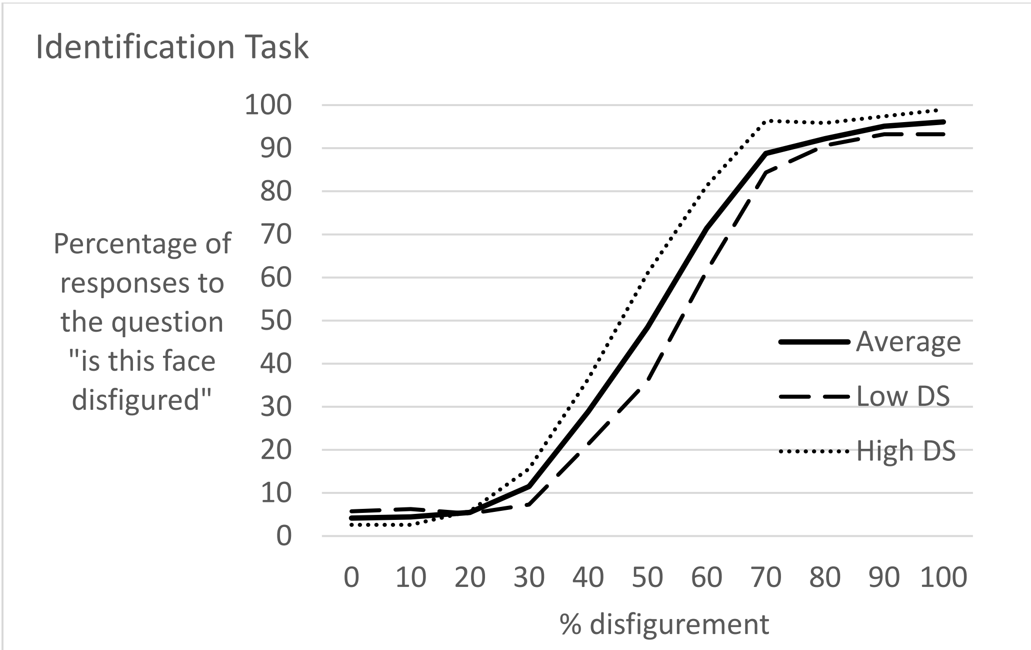


Figure 2. Experiment 1: Identification task: percentage of "yes" responses to the question "is this face disfigured" for each target individual and the average of all targets. ABX task: accuracy (left axis) and speed (right axis) of response in the discrimination task.

Figure 3: examples of stimuli used in Experiment 2.



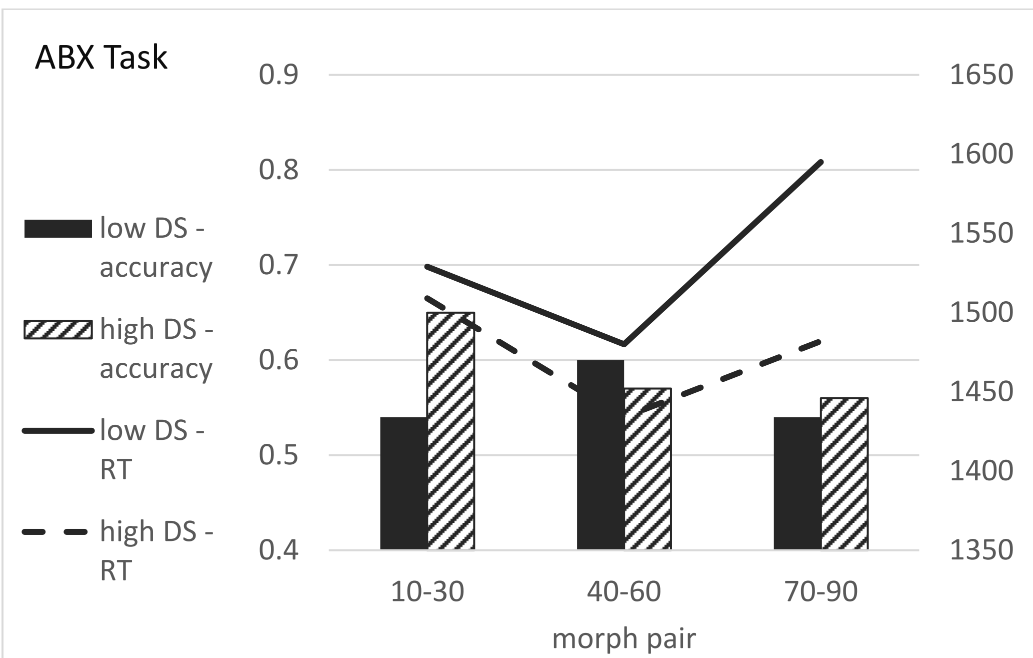


Figure 4. Experiment 2: Identification task: percentage of "no" and "yes" responses to the question "is this face disfigured". ABX task: accuracy (left axis) and speed (right axis) of responses to faces for participants with low or high disgust sensitivity.