THE UNIVERSITY OF CHICAGO

Examining the Memorability of Faces through Face Features

By

Daniel Buonauro

April 2022

A paper submitted in partial fulfillment of the requirements for the Master of Arts degree in the Master of Arts Program in the Social Sciences

Faculty Advisor: Daniel Albohn

Preceptor: Danielle Bolling

Abstract

Face and feature memorability serve an important role in understanding the complexities of human social interaction. The present study explored which facial features are most recognizable when a target face is rapidly presented for 200 ms. Exploratory analyses aimed to establish whether the facial features from target faces with the highest accuracy in a feature recognition task (Study 1A) were also the features most salient to participants in a check-box description task (Study 1B). Results from the recognition task indicated that participant accuracy across all features examined was significantly better than chance, and was highest for eyes, mouth, and eyebrows. Correlational results comparing the relationship between recognition accuracy and the saliency of features in a description task demonstrated that the relationship between the two is loosely connected. Taken together, these results provide a foundational framework from which to expand the understanding of feature recognition.

Examining the Memorability of Faces through Face Features

The faces encountered during everyday life leave different impressions on us: some are remembered at first glance, while others are forgotten. New faces are constantly introduced through social networks, in the media, and in person. From only a single exposure, some of these faces will remain in our minds, while other will fade. What allows some faces to last, while others quickly fade from memory?

Faces convey a wealth of information that is crucial to navigating everyday life. Information gathered from the face aids in social interaction, and can help identify critical personal information, such as identities and emotions (Curio et al., 2011). The importance of faces is not limited to social interaction. The media focuses our attention on faces, as they cover our television and movie screens, and are a consistent feature of both ancient and modern artwork. Even face-like patterns are often identified in everyday shapes and objects, such as in clouds and wall outlets, a phenomenon known as *pareidolia* (Palmer & Clifford, 2020). Within this context, face and facial feature memorability serve an important role in understanding the complexities of human interaction.

The encoding process of face stimuli has been the subject of much debate. Research in this area has typically focused on holistic and configural processing of face stimuli. This work has argued that faces are encoded as "wholes", often utilizing the spatial configuration of faces to do so. Each face contains spatial or configural properties that allows for it to be perceived as such. For example, the basic configuration of a face entails the eyes being above the nose, and the nose being above the mouth. This configural information also allows for the encoding of the spacing between, and positioning of, facial features. It has been argued that this configural-relational information is paramount in creating holistic memory representations of faces. The presence of holistic processing of faces has been demonstrated in a number of alternative ways. It has been argued that holistic processing is the dominant theory of face processing because individual face features, or piecemeal features (e.g. nose, mouth), are not represented in memory codes (Tanaka & Farah, 1993). Other work demonstrated that individual features are explicitly represented in memory codes, but are relatively inaccessible to conscious analysis or verbal report (Carey & Diamond, 1994). Further work has argued that piecemeal features are consciously accessible, but are internally encoded in a way that is influenced by other facial features (Bruce & Humphreys, 1994). The respective conclusions of this section of the literature differ, but they each stem from the common argument and observation that individual face features of upright face stimuli are typically difficult for individuals to assess singularly. For example, Young et al. (1987) created facial composites consisting of the top and bottom portions of two different famous faces. Non-composite stimuli were also created, which consisted of the top and bottom portions of the same face misaligned laterally. Participants were asked to identify either the top or bottom portion of each face. Results indicated that responses were slower when presented with composite stimuli than non-composite stimuli, but this effect disappeared when the face images were inverted. Young et al. argued that the upright composite stimuli were processed holistically, while the inverted composite stimuli were not. Later studies reinforced this conclusion that standard, upright face images were processed holistically, while inverted or fragmented faces, as well as non-facial stimuli, generally are not (Carey & Diamond, 1994; Endo et al., 1989).

There is converging evidence that upright face images evoke holistic processing in a second, similar line of research that has indicated spatial or ‘configural’ information is more readily encoded from upright face images rather than inverted faces (Bartlett & Searcy, 1993;

Leder & Bruce, 2000). Other configural research has demonstrated the capacity for aspects of both configural and featural processing. Searcy and Bartlett (2006) manipulated ordinary face stimuli in order to make them “grotesque”. This was done through two “distortions”: spatial and component. Spatial distortion was achieved either by moving facial features to slightly different locations on the face, such as moving the eyes slightly farther apart, or moving the mouth farther down the face. Component distortion was achieved by altering specific aspects of facial features, such as blackening several teeth. Participants rated the “grotesqueness” of each face, altered and unaltered, on a 7-point scale. They viewed and rated each face once while presented upright, and again while the images were inverted. Participants were then given a simultaneous paired-comparisons task in which they were presented with either a distorted face (spatial or component distortion) paired with an unaltered image of the same face (“different pairs”), or two identical unaltered faces (“same pairs”). Their task was to indicate whether the members of a pair of faces were identical or different. Results from the rating task indicated that inverting a face substantially impaired the encoding of spatial-configural information. Specifically, faces that were distorted through moving their eyes and mouths were judged as less “grotesque” when inverted than when upright. In contrast, faces that were distorted through changing the appearance of these same components were judged as approximately equally “grotesque” when presented upside down. These results provide evidence for a dual-mode view of face processing, which holds that spatial information and component information are encoded through separate processing modes. Results from the comparison task demonstrated that in the spatial distortion face trials, participants were able to correctly identify the items as different when presentation was upright. However, participants’ performance was significantly worse when presentation was inverted. Analysis of participants’ response times in the comparison task provided further evidence for the dual-mode position. Stimulus inversion significantly increased the response time in which participants responded to face images with features that were spatially different from the original image, but not those with altered features. This work not only demonstrated an effect of inversion on encoding of spatial aspects, but that the processing of spatial-configural information involves different operations than the processing of component information, highlighting the existence of feature-based face processing.

Configural processing is often contrasted with piecemeal, featural processing, which argues that individual facial features are crucial to the encoding of faces individually, and not because of their presence in constructing "whole" representations. It has been suggested in previous literature that individual facial features play an important role in face recognition (Leder & Bruce, 2000; Tversky & Krantz, 1969). Piecemeal features refer to aspects of faces that can be measured or described independently of one another, are local in their spatial context, and are marked by discontinuities in the surface of a face (Bartlett et al., 2003). A variety of evidence has suggested that the mechanisms involved in piecemeal face processing primarily analyze the internal features of the face image, such as the eyes, nose, and mouth (Maurer et al., 2002, Valentine, 1991). The ability to make use of differences in internal features tends to improve as faces become more familiar (Hancock et al., 2000). Ellis et al. (1979) found that internal (eyes, nose, mouth) and external (face shape, chin, ears) regions of faces can differ in their salience and effectiveness as cues in face recognition, and that perceivers make use of both of these regions when viewing unfamiliar faces, though internal features become more salient when recognizing familiar faces. Andrews et al. (2010) examined how these internal and external regions are represented in face-selective regions in the brain. Results indicated that there was a significant neural response in the fusiform face area to both the internal and external features of the face when each was presented in isolation, though the response was greater for internal features. Moscovitch et al. (1997) argued that the internal and external regions of the face may be processed differently in the brain. They concluded this based on their studies of CK, a victim of closed-head injury. CK was impaired at both lexical and object recognition, but performed normally in face recognition tasks if the face stimuli were presented upright. When face stimuli were inverted, CK performed significantly worse than normal participants. Moscovitch and colleagues compared CK to a control group in a face recognition task of well-known faces in two conditions. In one condition, the external region of the face stimuli was inverted, while the interior portion of the face remained upright. In the other condition, the external region remained upright, and the interior portion was inverted. When the external region of the face stimuli was inverted, results showed that recognition performance was unimpaired. However, when the internal region of the face stimuli was inverted, recognition performance was significantly disrupted, most especially with CK. Moscovitch and colleagues concluded that CK formed representations of faces based on orientation-specific configurations that consist primarily of internal features, further demonstrating the importance of the internal regions of the face in piecemeal encoding processes.

Previous research from large visual memory studies have shown that individuals have a remarkable ability to not only remember vast quantities of images, but specific image details as well (Brady et al., 2008; Vogt & Magnussen, 2007). This work has demonstrated that not only do observers have the capacity to remember thousands of images, but that these memories also contain detail. Further research determined that some images of scene images were consistently more forgettable than others, and that the memorability of these images can be predicted, highlighting the existence of intrinsic memorability properties of scene images (Isola et al., 2011). These memorability differences can be registered as quickly as 13 ms from the onset of a stimulus, and can last longer than a week (Broers et al., 2018; Goetschalckx & Wagemans, 2017).

Intrinsic memorability has also been observed in face images. Bainbridge et al. (2013) found that face memorability is an intrinsic feature of certain face photographs, suggesting that some faces are more likely to be remembered by the majority of people. Despite evidence for the intrinsic memorability of some faces over others, the subjective experience of memory is also substantially important. This has been demonstrated in research regarding own-race bias, whereby faces of one's own race are recognized more accurately than faces of other races (Chiroro & Valentine, 1995). This research has demonstrated that though intrinsic memorability can be predicted across a wide range of individuals, subjective experience has the capacity to influence the extent of this memorability.

The capacity for detail in long-term memory and intrinsic face memorability has been demonstrated using highly familiar faces. Ge et al. (2003) presented one well-known portrait of Mao Tse Tung, who is particularly familiar to their Chinese participants. The original target face was altered on the single dimension of interocular distance, which was either increased or decreased. Results displayed a highly accurate recognition of the exact face from the portrait, meaning participants were able to reproduce interocular distance, and recognize when it had been altered. This procedure was replicated by Bredart and Devue (2006) who generalized Ge et al.'s findings using face images of personally known individuals. Bredart and Devue presented individuals an unmanipulated target face of a close personal colleague mixed with interocular variations of the same face image. This research demonstrates the capacity for detail in long-term memory representations. Participants have the ability to utilize the spatial-configural information present in faces to accurately recognize the minute differences in the particular features of a face. Though spatial-configural information is traditionally considered holistic processing, the ability to recognize minute differences also points to the existence of featural representations in face processing.

Other research has sought to examine the underlying factors that influence intrinsic face image memorability. Sarno and Alley (1997) observed in a recognition task that faces rated as "distinct" were remembered at a higher rate than faces rated as "attractive". Attractiveness was a poor predictor of recognition, indicating that facial distinctiveness is a fundamental component in facial recognition performance. Wickham and Morris (2003) further divided distinctiveness into two measures: ease of spotting the face in a crowd, called the "traditional" measure, and deviation from the average face, called "deviation". Results showed that when the effects of age were removed, attractiveness did not have a significant relationship with the recognition of the face. However, the traditional measure of distinctiveness was found to be closely related to memorability, indicating that faces more likely to be spotted in a crowd will also be remembered more readily in a recognition task.

Current Study

Previous literature on the encoding processes of faces has typically focused on the holistic and configural aspects of the face in order to gain insight into the mechanisms that underlie face perception. Though some work has demonstrated the prevalence of both holistic and piecemeal mechanisms of processing, it is not fully understood how individual face features may contribute to the memorability of a target face after a single, brief exposure. Previous research in featural processing has shown a clear importance of internal features in the perception and recognition of faces (Moscovitch et al., 1997). However, research in holistic processing provides evidence that internal and external features of a face are combined to create an overall impression (Carey & Diamond, 1994). What remains to be examined is the relative importance of specific face features over others in a quickly-paced recognition task, most especially when the face stimuli are upright. Other work has demonstrated the vast long-term memory capacity for detailed images, and the intrinsic memorability of face images (Bainbridge et al., 2013; Brady et al., 2008; Vogt & Magnussen, 2007). However, it is not understood whether specific facial features are accounted for in these memory representations, and if they are, whether some features are more strongly represented than others. The current study aims to explore this gap in the literature by examining the memorability of target faces through a focus on the individual facial features. If featural encoding processes exist as described in previous literature, then participants should be able to recognize the specific facial features of face stimuli at a rate better than chance.

The present study aims to conduct exploratory analyses in order to examine the role of individual features in face recognition memory. Experiment 1A will examine whether participants are able to recognize the specific facial features of target faces at a rate higher than chance in a recognition task. Experiment 1B will examine which features are most salient to participants in a check-box description task. Experiment 1B will also examine whether the features most often identified and selected for each target face in the check-box task are also the features with the highest accuracy in the same target faces in the recognition task in experiment 1A. This serves to connect the task performance of experiment 1A with the intuitive feature identification and salience of experiment 1B. In line with previous work, I predict that the highest recognition accuracy will be demonstrated for the internal features of target faces. Specifically, I predict that eye recognition trials, followed by mouth, eyebrows, and nose trials, will exhibit the highest recognition accuracy, compared with external facial features such as face shape, ears, and chin. I also predict a significant positive correlation between performance in the recognition task and the proportion of responses of that particular feature in the check-box description task. In other words, I predict that the features most often described for a particular target face will positively correlate with recognition accuracy for those same features. This relationship would thus constitute a theoretical link between the facial features participants actually remember, as demonstrated in the recognition task, and the features intuitively most noticeable to participants, as demonstrated in the check-box descriptor task.

Study 1A

Experiment 1A aimed to examine the role of individual features in face recognition by assessing whether participants are able to recognize the specific facial features of target faces at a rate higher than chance (50% accuracy). In other words, experiment 1A examined whether participants will be able to recognize the eyes of a target face when compared to another, similar set of eyes. Furthermore, this experiment aimed to establish which features will be recognized with higher accuracy than other features.

Methods

Participants

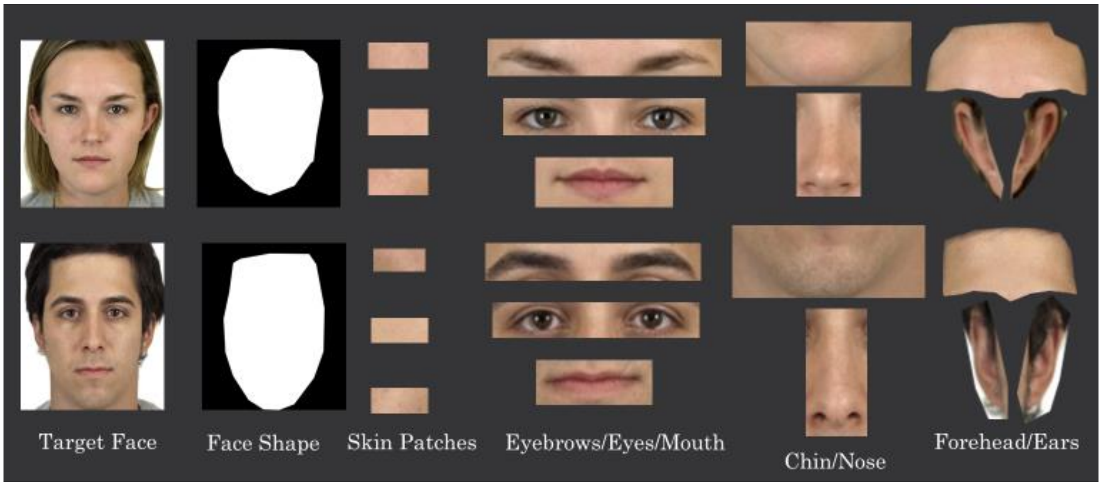
168 participants (106 male, 62 female) were recruited remotely through CloudResearch, a participant-sourcing platform for online research. 131 participants were Caucasian, 14 were African-American, 14 were Asian, and 5 identified as "Other." Participants were 18-55 years of age (M = 37.45). Participants were compensated for their time.

Materials

Target faces were gathered from the Chicago Face Database, an online face image database (Ma et al., 2015). Sixty Caucasian face images were randomly selected from the database for use (30 male and 30 female faces). The faces were similar in age, and had neutral facial expressions. Face images were manually cropped using GIMP, an image manipulation program, to create facial features for test. The features included face shape, forehead, eyes, eyebrows, ears, nose, mouth, chin, and skin patches from the left and right cheeks as well as the center of the forehead. Sixty trials out of a possible 660 (11 features x 60 target faces) were randomly selected for each participant. See Figure 1 for an example of face features extracted from one female and one male target face.

Figure 1

Examples of Features Extracted from One Male and One Female Target Face

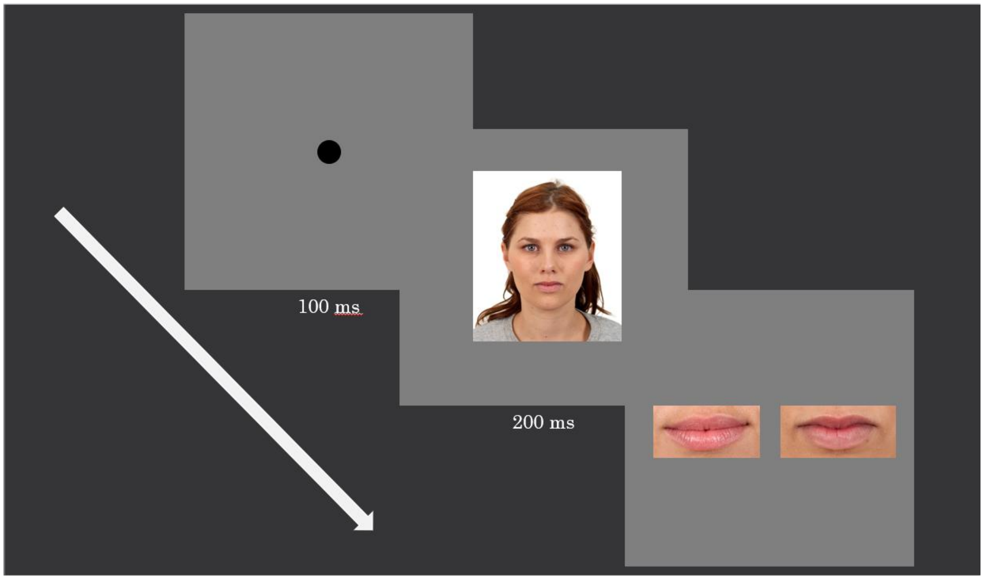


Procedure

After agreeing to participate in the study on CloudResearch, participants were instructed to select the feature that belonged to the previous face. Participants' attention was first drawn to the center of the screen via a fixation dot. A target face stimulus was then presented on screen for 200 ms before being immediately replaced with two similar face feature snippets, one from the target face and one from a different, distractor face side by side. The distractor features were taken from one of the other 60 faces in the stimulus set but always matched the target feature (e.g., mouths were always paired with mouths). It was not indicated to the participant which feature was to be remembered in any trial. Participants were asked to select which of the presented features belonged to the previous face. Responses were made using keyboard responses that required the use of the index finger on both hands. Participants completed each trial at their own pace, and had no time restrictions between trials. See Figure 2 for task design.

Figure 2

Example of Trial Layout – Study 1A



Data Preparation and Analysis Strategy

All analyses were conducted with a confidence interval of 0.95. Responses were assessed at the participant level by first calculating the proportion correct for all feature trials across all participants. This way, overall performance in the task could be assessed. Similarly, this approach allowed for determining whether or not participants were able to accurately recognize features better than chance. The proportion of correct recognition responses by each participant for each feature tested was then calculated. This was done by dividing the total number of trials seen in each feature conditions by the number of correct responses for each feature condition. For example, if a participant viewed 10 mouth trials, and correctly recognized 8 mouths, their proportion was 80%, or .8. Participants thus had a proportion, or percentage correct, for each respective feature across all trials seen.

A series of one-sample t-tests were conducted to calculate whether, at the participant level, each feature examined was correctly recognized at a rate higher than chance, meaning 50% accuracy. A repeated measures ANOVA was run using feature as the within-subjects variable in order to examine whether there was a significant main effect of feature condition on accuracy. Pairwise comparisons were then conducted to examine whether the 55 feature pairings were significantly different from one another. Tukey's method for comparing a family of 11 estimates was used for p-value adjustment.

Results

Results indicated that, overall, participants were significantly better than chance at recognizing features from target faces (M = 0.63, SD = 0.08, $t(167)$ = 22.80, p < .01). That is, participants were able to correctly recognize the feature that belonged to the target face in 63% of all trials. A series of one-sample *t*-tests were conducted to calculate whether each feature examined was correctly recognized at a rate higher than chance, meaning 50% accuracy (See Table 1).

Table 1

Results from One-Sample t-Tests

Feature	Mean	*SD*	*t*	df	*p*-value	Cohen's *d*
Face Shape	0.56	0.19	4.82	167	< .001	0.34
Forehead	0.61	0.18	8.97	167	< .001	0.63
Eyebrows	0.68	0.20	12.71	167	< .001	0.89

Eyes	0.73	0.19	16.98	167	< .001	1.19
Nose	0.61	0.18	8.09	167	< .001	0.56
Ears	0.61	0.19	7.45	167	< .001	0.52
Mouth	0.70	0.19	14.31	167	< .001	0.99
Chin	0.66	0.20	11.18	167	< .001	0.78
Left Skin Patch	0.60	0.18	8.39	167	< .001	0.59
Center Skin Patch	0.59	0.19	7.08	167	< .001	0.49
Right Skin Patch	0.60	0.19	7.70	167	< .001	0.54

Note. Means presented in this table presented as percentage of trials correct.

As indicated in Table 1, participants were able to correctly recognize each facial feature at a rate significantly higher than chance.

The repeated measures ANOVA revealed a significant main effect of feature on recognition accuracy, $F(10, 167) = 15.99$, $p < .001$. Pairwise comparisons on all 55 pairings revealed that performance in the recognition task was particularly significant in eye, mouth, and eyebrow trials. Participants were most accurate in eye recognition trials ($M = .73$, $SD = .19$). Accuracy in eye recognition was significantly higher when compared to all other features except mouths ($M = .70$, $SD = .19$) and eyebrows ($M = .68$, $SD = .20$). Comparisons revealed that eye recognition was significantly more accurate than face shape recognition, $t(167) = 8.51$, $p < .001$. Eye recognition was also significantly more accurate than recognition of each skin patch, left: $t(167) = 6.82$, $p < .001$, right: $t(167) = 6.87$, $p < .001$, center: $t(167) = 6.82$, $p < .001$. Eye recognition was significantly more accurate than ear, $t(167) = 7.68$, $p < .001$, and nose recognition, $t(167) = 7.66$, $p < .001$. The same was true for forehead recognition, $t(167) = 6.78$, $p < .001$, and chin recognition, $t(167) = 3.92$, $p = .005$. Eye recognition was marginally different from eyebrow recognition, $t(167) = 3.16$, $p = .07$. Eye recognition was not significantly different from mouth recognition, $t(167) = 1.87$, $p < .73$.

Comparisons also revealed that mouth recognition was significantly more accurate than face shape recognition, $t(167) = 6.77$, $p < .001$. Mouth recognition was significantly more accurate than recognition of each skin patch: left, $t(167) = 5.44$, $p < .001$, right, $t(167) = 4.95$, $p < .001$, and center, $t(167) = 5.77$, $p < .001$. Mouth recognition was also significantly more accurate than ear, $t(167) = 5.66$, $p < .001$, and nose recognition, $t(167) = 6.02$, $p < .001$, as well as forehead recognition, $t(167) = 5.04$, $p < .001$. Notably, mouth and chin recognition were not significantly different from one another, $t(167) = 2.33$, $p = .41$.

Comparisons further revealed that eyebrow recognition was significantly more accurate than face shape recognition, $t(167) = 5.93$, $p < .001$. Eyebrow recognition was also significantly more accurate than recognition of each skin patch: left, $t(167) = 3.83$, $p = .007$, right, $t(167) = 3.78$, $p = .009$, and center, $t(167) = 4.22$, $p = .001$. Eye recognition was significantly more accurate than ear $t(167) = 4.19$, $p = .002$ and nose recognition $t(167) = 4.29$, $p = .001$, as well as forehead $t(167) = 3.59$, $p = .01$. Eyebrow recognition was not significantly different from chin recognition, $t(167) = 1.02$, $p = .99$.

Other comparisons revealed that chin recognition was significantly more accurate than face shape, $t(167) = 4.89$, $p < .001$, and nose recognition, $t(167) = 3.53$, $p = .02$. Chin recognition was also significantly more accurate than center skin patch, $t(167) = 3.42$, $p = .03$. However, chin recognition was not significantly different from either right skin patch recognition, $t(167) = 2.87$, $p = .13$, or left skin patch recognition, $t(167) = 2.93$, $p = .12$.

Discussion

Experiment 1A examined participants' ability to recognize specific features from target face images. As predicted, results from a feature recognition task indicated that participants were able to recognize each facial feature significantly better than chance, though some features exhibited a higher degree of recognition accuracy than others. Eyes, mouth, and eyebrow trials yielded the highest recognition accuracy, respectively. These results are consistent with eye-tracking research in dyadic conversations. Rogers et al. (2018) utilized eye-tracking to examine personal gaze patterns during social interaction. Rogers and colleagues found that personal eye gaze patterns of individuals engaged in face-to-face conversation exist in an eye-mouth gaze continuum, in which some participants exhibited a strong preference for eye gaze, some exhibited a strong preference for mouth gaze, and others distributed their gaze between the eyes and mouth to varying extents. This work may provide insight into preferential encoding and subsequent increased accuracy of recognition of specific facial features that are involved most heavily in communication.

The eyes, eyebrows, and mouth have been demonstrated to be crucial in providing information about the emotional states of others (Baron-Cohen et al., 1997; Smith et al., 2005; Wegrzyn et al., 2017). For example, Wegrzyn et al. (2017) found that participants heavily relied on the eye and mouth regions over all other features when successfully recognizing an emotion in a target face. The ability to accurately interpret facial expressions is of primary importance for humans to socially interact with one another (Nachson, 1995). This social importance may be a mechanism for which preferential feature encoding is enabled for eye and mouth regions.

The biases toward the eye and mouth regions during conversation paired with the importance of these features (in addition to the eyebrows) in emotion recognition provide possible explanations for why nose recognition was not as accurate as the other internal features of target faces. These results thus only provide partial support for featural processing research that highlights the importance of internal regions of the face (Ellis et al., 1979; Moscovitch et al., 1997). Though participants were able to recognize each internal region of the face significantly better than chance (eyebrows, eyes, nose, mouth), nose recognition was not as strong as the other internal features, which would be expected if feature recognition accuracy was solely a matter of encoding preference for the internal region of faces.

Study 1B

Study 1B aimed to examine which features were most often noticed and mentioned by participants when presented with target faces in a check-box description task. Combined with the results from Study 1A, Study 1B was conducted to determine if successful face feature recognition is the result of some features being subjectively more noticeable than others on specific target faces.

Methods

Participants

99 participants (57 male, 42 female) were recruited remotely through CloudResearch, a participant-sourcing platform for online research. 73 participants were Caucasian, 15 were African-American, 10 were Asian, and 1 identified as “Other.” Participants were 22-76 years of age (M = 39.01). Participants were compensated for their time.

Materials

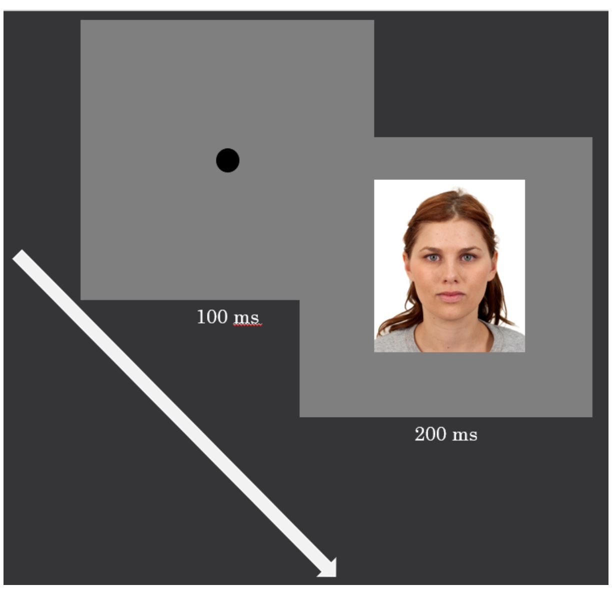
The same 60 faces from the CFD used in Study 1A were used as stimuli in Study 1B, though not manipulated. Participants randomly received a selection of 10 face images of the 60 possible images. Check-box descriptions were developed with the intent to highlight and differentiate the specific aspects of each facial feature examined. For example, each participant was presented with descriptors such as: “Ears: Pronounced/Pointed” and “Ears: Flat.” See Appendix A for all check-box descriptors used. Each participant received the same set of descriptors for each target face.

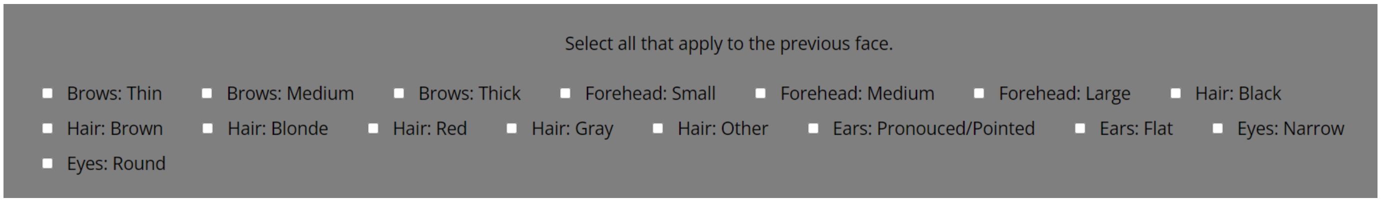
Procedure

After agreeing to participate in the study on CloudResearch, participants were instructed to check the box beside the feature descriptors that best applied to the presented target face. Participants were not required to select a descriptor related to each feature. Participants had no minimum or maximum number of check-boxes that needed to be filled in order to complete the task. They were only asked to select all that apply. Similar to the design of experiment 1A, participants' attention was first drawn to the center of the screen via a fixation dot. Then, a target face stimulus was presented on screen for 200 ms before being immediately replaced with the check-box descriptors. The descriptors remained on screen until the participant advanced. Participants completed each trial at their own pace, and had no time restrictions between trials. Responses were made by using the computer cursor to select the box beside each descriptor on screen. See Figure 3 for task design and example of check-box descriptors.

Figure 3

Example of Trial Layout – Study 1B





Note. Not all check-box descriptors shown. See Appendix A for full list of descriptors utilized.

Data Preparation Analysis Strategy

Responses were first analyzed at the participant level to determine whether each participant checked a box related to a specific feature (e.g., “Ears: Flat” related to ears). These response counts were then averaged at the target face level, in order to identify the extent to which each descriptor was selected for a particular target face. Z-scores were computed to analyze which features differed most from the average. Z-scores were similarly computed from the data of Study 1A, which measured the proportion correct of each feature of each target face. This was calculated as the proportion correct for each feature of each target face across all participants. Each of the 60 target faces thus had a proportion correct for each feature (60 target faces x 11 features = 660 total proportions of correct responses). Thus, 11 Pearson’s *r* correlations were conducted between the two sets of z-scores in order to analyze whether the features of target faces most often recognized in Study 1A were correlated with the features most often identified in Study 1B. Each correlation describes the relationship between the z-score of the proportion correct of one type of feature of a target face and the z-score proportion of the mentions of that feature in the check-box description task. All correlations were conducted with a significance level of .05.

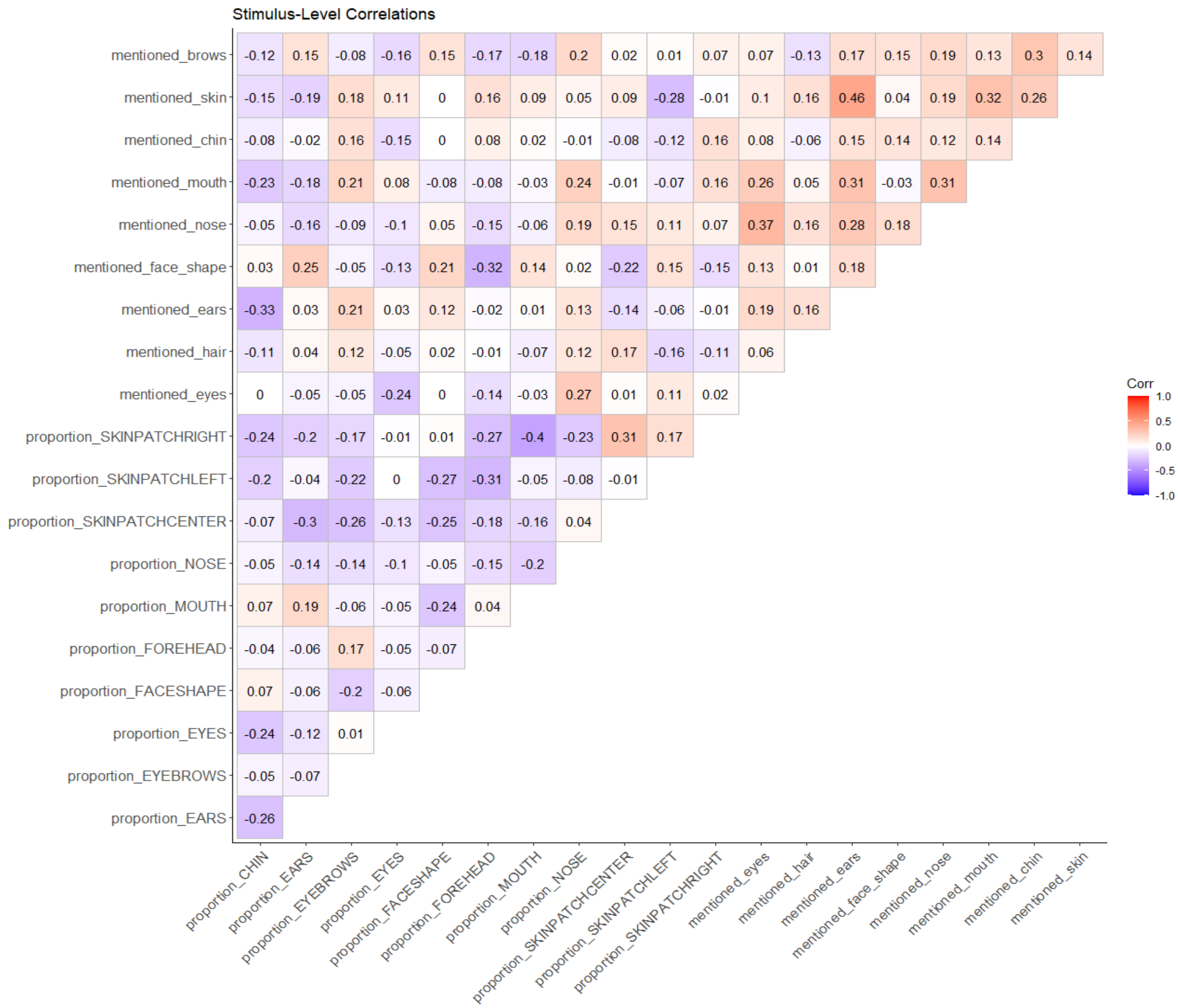
Results

Only one correlation was significant at the .05 level. The correlation between the proportion correct of left skin patch recognition responses compared with the check-box descriptor prevalence of skin mentions was significant, $r(53) = -0.28$, $p = .03$. That is, the higher the proportion of check-boxes selected that mentioned skin features of a target face, the worse the performance was for left skin patch in recognition trials. However, this relationship was not maintained in right skin patch, $r(53) = -.008$, $p = .95$, or center skin patch, $r(53) = -0.09$, $p = .50$. The correlation between the proportion correct of eye trials compared with the check-box descriptor prevalence of eye mentions was marginally significant, $r(53) = -.24$, $p = .08$. That is, the higher proportion of check-boxes selected that mentioned the eyes of a target face, the worse the performance was for eyes in recognition trials. The other features examined that exhibited the highest accuracy in the recognition task (experiment 1A), mouth, $r(53) = -0.03$, $p = .82$, and eyebrows, $r(53) = -.08$, $p = .55$, did not significantly correlate with the proportion of mentions for each respective feature. Though considered an internal feature, the correlation between the proportion correct of nose responses compared with the check-box descriptor prevalence of nose mentions was also not significant, $r(53) = .19$, $p = .17$.

External features demonstrated similar patterns to internal ones. The correlation between the proportion correct of ear responses compared with the check-box descriptor prevalence of ear mentions was not significant, $r(53) = .02$, $p = .84$. The correlation between the proportion correct of face shape responses compared with the check-box descriptor prevalence of face shape mentions was also not significant, $r(53) = .21$, $p = .12$. Finally, the correlation between the proportion correct of chin responses compared with the check-box descriptor prevalence of chin mentions was not significant, $r(53) = -.34$, $p = .56$.

Figure 4

Correlation Matrix of Relationship between Mentions and Accuracy



Note. The correlations examined in the figure above are the ones that match specific feature mentions with that same feature's accuracy. For example, mentioned mouth and proportion mouth.

Discussion

Experiment 1B aimed to assess whether there is a significant relationship between the facial features of target faces that are most identifiable in a description task and the recognition accuracy of those same features. This allowed for insight into whether the features participants "think" they might remember, indicated by their ability to describe that feature using the provided descriptors, will correlate with what they “actually” remember, indicated by the recognition accuracy of those same features in the recognition task (Study 1A). A target face was presented on screen for 200 ms before being replaced with a series of check-box descriptors. Participants checked the box beside the descriptors of features that best applied to a particular target face. Though there was one significant correlation, left skin patch, it was in the opposite direction of initial predictions. In other words, the higher the proportion of check-box descriptors selected that mentioned the skin of a target face, the less accurate the left skin patch was in the recognition task. This same pattern did not achieve significance in any other feature examined, including the other two skin patches. This makes it difficult to form concrete conclusions based on the single significant correlation. The correlation between the proportion of check-box descriptors selected that mentioned the eyes of a target face and the recognition accuracy of those same eyes was marginally significant, though it exhibited a negative relationship. That is, the higher the proportion of check-box descriptors selected that mentioned the eyes of a target face, the less accurate eye recognition was. No other relationship examined was significant or approached significance in either direction.

This study provided insight as to whether or not the features that are most salient to participants when asked to describe the target face are the same features that are most likely to be accurately selected in the recognition task. The present study demonstrated that the relationship is loosely connected, though further research is necessary to more concretely determine this connection.

General Discussion

The current study aimed to explore feature recognition performance, and whether the selections made with regard to specific features in a check-box description task correlated with the recognition accuracy of those same features. This work can be viewed as an attempt to draw a connection from the 'intuitive' check-box description task, meaning what features stood out most to participants, to the 'actual' behavioral performance in a feature recognition task.

In line with predictions, results from the recognition task (Study 1A) showed that for all 11 features tested, participants were able to recognize which feature belonged to a presented target face at a rate significantly higher than chance. Though each feature was found to be recognized significantly better than chance, some features were more likely to be recognized than others. The eyes were the most consistently recognized face feature across all participants, followed by mouths, and eyebrows. These results are consistent with research that has examined the gaze of participants engaged in one-on-one conversation. Results from these studies have shown that participants tend to fixate their gaze on the other individuals' eyes, mouth, or alternating back and forth between the two (Rogers et al., 2018). The results from the present study could be attributed to the increased selective attention given to these features in a conversational setting, creating a learned disposition for attending to those same features when presented with faces in other settings.

The eyes and mouth have been found to be the most crucial in recognizing emotion (Baron-Cohen et al., 1997; Wegrzyn et al., 2017). Baron-Cohen et al. (1997) demonstrated that adults have shown to be remarkably consistent in how emotions are interpreted in faces, for both basic (happy, sad, angry) and complex (admire, interest, thoughtfulness) mental states. The eyes in particular were found to convey as much information as the whole face in judgments about complex mental states. These results provide support for the "language of the eyes", a nonverbal communicative channel that allows for detecting emotional states of others solely from the information provided by the eyes (Baron-Cohen, 1995).

The results of the present study provide partial support for featural processing research that highlights the importance of internal regions of the face (Ellis et al., 1979; Moscovitch et al., 1997). Participants were able to recognize each internal feature of the face better than chance (eyebrows, eyes, nose, mouth). Recognition accuracy for the eyes, mouth, and eyebrows were not significantly different from each other. However, recognition accuracy for each of those three features was significantly higher than the recognition accuracy in nose trials. Nose recognition was not as strong as the other internal features, which would be expected if feature recognition accuracy was solely a matter of encoding preference for the internal region of faces. As recognition accuracy was also significant for external features of the face (face shape, ears, chin, skin patches), further research is needed to determine the extent of the encoding differences between internal and external features of the face.

Study 1B aimed to assess which facial features are most noticeable to participants in a check-box description task, and whether the features most often selected in this description task would correlate with the behavioral performance of the recognition task. For example, if a target face had a high proportion of check-box selections that mentioned a particular feature, this might provide insight into the "intuitive memorability" – or what individuals *think* they will remember– of that particular face in a recognition task. The relationship between feature mentions and feature accuracy was not strongly supported in the present study. There was a negative significant relationship between skin mentions and left skin patch recognition accuracy, center and right skin patch were not significant. This does not allow for concrete conclusions to be made about the nature of the relationship between skin mentions and recognition accuracy. There was also a marginally significant negative relationship between the proportion of eye mentions and the proportion of eye responses in recognition trials. Though this relationship was approaching significance, it was in the opposite direction as predicted. While only two of the 11 correlations reached significance or marginal significance, this relationship deserves further speculation. It is worth noting that the pattern of the relationships for both the eyes and skin was negative, suggesting that participants overestimate the salience of some features in comparison to others. Further research is needed to more concretely determine the nature of the relationship between the features with the highest recognition accuracy and the salience of those same features in a description task.

These exploratory analyses provide an important foundation in which to explore further feature-based examinations. One such direction is to combine studies 1A and B into a single within-subjects design. This would enable a more direct examination and connection of whether the features most accurately recognized in study 1A are the same as the features most often identified in study 1B for the same participants. For example, one individual may think they will remember the eyes of a target face while another believes they will remember the mouth. Recognition accuracy for those specific features can then be examined in order to more concretely determine the relationship between what features participants *think* they will remember from a target face and what they *actually* remember in a recognition task.

Another possible future direction for this research would be to examine the effects of minimal group assignment on face feature memory. In other words, this research would determine whether participants better recognize the features of those in their ingroup compared with their outgroup. Tajfel et al. (1971) assessed the effects of social categorization on intergroup behavior using minimally different information between groups, concluding that participants would act in ways to benefit their ingroup, often at the expense of the outgroup, and in spite of the arbitrary nature of the group division. Recent research in intergroup processing has shown that meaningless group labels can create bias in mental representations of faces, evaluations, and trait inferences, and can imply characteristics about groups that may influence responses in intergroup contexts (Hong & Ratner, 2021). If an ingroup encoding bias exists in face processing, perhaps this bias would also be represented at the featural level.

A limitation of the current study is the race of the target faces selected for use was limited to Caucasian target faces. Expanding the race of target faces could provide insight into implicit racial bias in face recognition, such that participants could be more likely to correctly recognize the features of own-race target faces. An additional limitation is the samples of experiments 1A and 1B were largely Caucasian and male. A more balanced sample could lead to analyses examining the potential interaction between the race of the participant and feature recognition performance, as well as the interaction between the race of the target face and feature recognition performance. The present study utilized target faces and features with neutral emotional expression. Other research has explored the association between perceived personality and emotionality in face memory. Mattarozzi et al. (2014) found that faces perceived as trustworthy and untrustworthy were both remembered more often than “neutral trustworthy” ones. This emotionality bias has been shown to extend to positive emotional affect as well. Gallegos and Tranel (2005) found that an emotional facial expression (happiness) helped facilitate the identification of familiar faces, though it is unclear whether more expressive features (e.g., eyes, eyebrows, mouths) indicate more memorability. Franklin and Adams (2010) found that reading another’s mental state from a face can cause a person to form deeper associations with that face and consequently remember it with higher accuracy. Emotional facilitation was also explored in compound emotional states. Laplante and Ambady (2000) demonstrated that the accuracy with which faces displaying compound expressions (sadness and anger) were recognized was significantly higher than the accuracy with which faces displaying singular expressions (sadness alone) were recognized. Further research is needed in order to more clearly understand the relationship between emotionality and memorability, and the role individual features such as the eyes and mouth play in predicting and recognizing the emotional states of others.

Conclusion

The present study demonstrated the memorability of facial features in a recognition task. These results point to the vast capacity for feature memory in face processing. As analyses were exploratory, the present study provided a foundational framework from which to expand the understanding of feature recognition of target faces. Further research is needed to explore the relationship between the features of specific target faces that are most noticeable in a description task and the memorability of those same features in a recognition task. If a target face had a high proportion of check-box descriptor selections that mentioned a particular feature, this might provide insight into the "intuitive memorability" – or what individuals *think* they will remember– of that particular face when later presented in a recognition task. The connection between the "intuitive" and the "actual" memorability of features of target faces deserves further study in order to provide a deeper understanding of face feature memorability.

References

Andrews, T.J., Davies-Thompson, J., Kingstone, A., & Young, A. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *Journal of Neuroscience, 30*(9), 3544-3552.

Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, *142*(4), 1323–1334.

Baron-Cohen, S. (1995). Mindblindness: An essay on autism and theory of mind. Cambridge, MA: MIT Press/Bradford Books

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger syndrome. *Vis Cogn*, *4*(3), 311–31.

Bartlett J.C., Searcy J.H., Abdi, H. (2003). What are the routes to face recognition? In M. Peterson, & G. Rhodes (Eds.): *Perception of Faces, Objects, and Scenes: Analytic and Holistic Processes.* Oxford: Oxford University Press. (pp. 21-52).

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325-14329.

Brédart, S., & Devue, C. (2006). The accuracy of memory for faces of personally known individuals. *Perception, 35*(1), 101–106.

Broers, N., Potter, M.C. & Nieuwenstein, M.R. (2018). Enhanced recognition of memorable pictures in ultra-fast RSVP. *Psychonomic Bulletin Review, 25*, 1080–1086.

Bruce, V. & Humphreys, G. W. (1994). Recognizing objects and faces. *Visual Cognition, 1*, 141-180.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: How do we tell the difference between male and female faces? *Perception*, *22*(2), 131–152.

Carey S., & Diamond R. (1994). Are faces perceived as configurations more by adults than by children? *Vis. Cogn.* 1, 253–274.

Chiroro, P., & Valentine, T. (1995). An investigation of the contact hypothesis of the own-race bias in face recognition. *The Quarterly Journal of Experimental Psychology, 48*(4), 879-894.

Curio, C., Bülthoff, H. H., & Giese, M. A. (Eds.). (2011). Dynamic faces: Insights from experiments and computation. *MIT Press.*

Ellis, H.D., Shepherd, J.W., Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features. *Perception, 8,* 431–439.

Endo, M., Masame, K., & Maruyama, K. (1989). Interference from configuration of a schematic face onto the recognition of its constituent parts. *Tohoku Psychologica Folia, 49,* 114-125.

Franklin, R. G., Jr., & Adams, R. B., Jr. (2010). What makes a face memorable? The relationship between face memory and emotional state reasoning. *Personality and Individual Differences*, *49*(1), 8–12.

Gallegos, D.R., & Tranel, D. (2005). Positive facial affect facilitates the identification of famous faces. *Brain and Language*, *93*(3), 338-348.

Ge, L., Luo, J., Nishimura, M., Lee, K. (2003). The lasting impression of Chairman Mao: hyperfidelity of familiar-face memory. *Perception*, *32*, 601–614.

Goetschalckx, L., & Wagemans, J. (2017). Towards a better understanding of image memorability. In *Cognitive Neuroscience of Memory: The Recollection, Familiarity, and Novelty Detection conference, Date: 2017/03/23-2017/03/24, Location: Liège, Belgium*.

Hancock, P.J., Bruce, V., Burton, A.M. (2000). Recognition of unfamiliar faces. *Trends Cogn Sci, 4*(9):330-337.

Hong, Y., Ratner, K. G. (2021). Minimal but not meaningless: Seemingly arbitrary category labels can imply more than group membership. *Journal of Personality and Social Psychology, 120*(3), 576-600.

Isola, P., Xiao, J. X., Torralba, A., & Oliva, A. (2011a). What makes an image memorable? *24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 145-152.

LaPlante, D., Ambady, N. (2000). Multiple messages: facial recognition advantage for compound expressions. *Journal of Nonverbal Behavior, 24*, 211–224.

Leder, H., & Bruce, V. (2000) When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology, 53A*(2), 513-536

Ma, Correll, & Wittenbrink (2015). The Chicago face database: a free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122-1135

Mattarozzi, K., Todorov, A., & Codispoti, M. (2015). Memory for faces: the effect of facial appearance and the context in which the face is encountered. *Psychological Research*, *79*(2), 308-317.

Maurer, D., Le Grand, R., Mondloch, C.J. (2002). The many faces of configural processing *Trends in Cognitive Science, 6,* 255–260.

Moscovitch, M., Wincour, 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*, 555-604.

Nachson, I. (1995). On the modularity of face recognition—the riddle of domain specificity. *Journal of Clinical and Experimental Neuropsychology*, *17*, 256–275.7629271.

Oliva, A., Isola, P., Khosla, A., & Bainbridge, W. A. (2013). What makes a picture memorable? *SPIE Newsroom Article, 7 May 2013*.

Palmer, C. J., & Clifford, C. W. G. (2020). Face pareidolia recruits mechanisms for detecting human social attention. *Psychological Science*, *31*(8), 1001–1012.

Rogers, S.L., Speelman, C.P., Guidetti, O. (2018) Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific Reports, 8,* 4271.

Sarno, J.A., & Alley, T.R. (1997). Attractiveness and the memorability of faces: Only a matter of distinctiveness? *American Journal of Psychology, 110*(1), 81-92.

Searcy, J. & Bartlett, J. C. (1996) Inversion and processing of component and spatial-relational information in faces. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 904-915.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology*, *75,* 221-242.

Smith, M. L., Cottrell, G. W., Gosselin, F., & Schyns, P. G. (2005). Transmitting and decoding facial expressions. *Psychological Science, 16*(3), 184–189.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*, 149 –178.

Tanaka, J.W., & Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology A,* 46, 225–245.

Tversky, B., & Krantz, D. H. (1969). Similarity of schematic faces: A test of interdimensional additivity. *Perception and Psychophysics, 5*, 124-128.

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.

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology, 54*(4), 298-303.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PLOS ONE, 12*(5).

Wickham, L. H. V. and Morris, P. E. (2003). Attractiveness, distinctiveness, and recognition of faces: Attractive faces can be typical or distinctive but are not better recognized. *American Journal of Psychology*, *116*(3), 455–468.

Young, A. L. (1991) Deformable templates for face recognition. *Journal of Cognitive Neuroscience, 3*, 59-70.

Young, A.W., Hellawell, D., & Hay, D.C. (1987). Configurational information in face perception. *Perception*, *16*(6):747-59.

Appendix A

Full list of descriptor items:

Eyes: Green

Eyes: Brown

Eyes: Blue

Eyes: Hazel

Eyes: Other

Eyes: Round

Eyes: Narrow

Eyes: Close

Eyes: Average

Eyes: Far

Hair: Black

Hair: Blonde

Hair: Red

Hair: Gray

Hair: Brown

Hair: Other

Hair: Straight

Hair: Wavy

Hair: Curly

Hair: Short

Hair: Medium

Hair: Long

Facial Hair: None

Facial Hair: Stubble

Facial Hair: Moustache

Facial Hair: Beard

Facial Hair: Other

Ears: Pronounced/Pointed

Ears: Flat

Forehead: Small

Forehead: Medium

Forehead: Large

Face: Short

Face: Medium

Face: Large

Face: Oval

Face: Round

Face: Angular

Face: Other

Face Symmetry: Symmetric

Face Symmetry: Average

Face Symmetry: Asymmetric

Nose: Small

Nose: Medium

Nose: Large

Nostrils: Small

Nostrils: Large

Lips: Thin

Lips: Medium

Lips: Thick/Full

Chin: Angular

Chin: Round

Skin Marks: Scar

Skin Marks: Acne

Skin Marks: None

Skin Marks: Wrinkles

Skin Marks: Other

Skin: Light

Skin: Medium

Skin: Dark

Cheeks: Low/Set back

Cheeks: Average

Cheeks: High/Arched

Brows: Thin

Brows: Medium

Brows: Thick