Effect of vertical stretching on the extraction of mean identity from faces

Article (Accepted Version)

Alexandra Bagaïni

Graham Hole*

School of Psychology,

University of Sussex,

Falmer,

Brighton,

BN1 9QH.

*Please address all correspondence to the second author, grahamh@sussex.ac.uk.

Abstract

Observers can extract the mean identity from a set of faces and falsely recognise it as a genuine set member (de Fockert & Wolfenstein, 2009). The current experiment demonstrated that this "perceptual averaging" also occurs with vertically stretched faces. On each trial, participants decided whether a target face was present in a preceding set of four faces. In the *control* condition, the faces were all normally proportioned; in the *stretched set* condition, the face sets were stretched but the targets were normal; and in the *stretched target* condition, the face sets were normal but the targets were stretched. In all three conditions, participants falsely identified the set mean as a face that had been presented within the set, implying that this identity-averaging effect is based on high-level identity information rather than the low-level physical characteristics of the face stimuli.

Keywords

Set averaging, ensemble coding, face recognition, vertical stretching

Introduction:

When presented with a set of similar stimuli, we often seem to create an "average" representation from them, a process variously described as "perceptual averaging", "set averaging" or "ensemble coding". Most studies of this have been based on Ariely's (2001) method. On each trial, he showed participants a set of circles that varied in size, followed by a target circle. If the target's size corresponded to the mean size of the preceding circles, participants tended to mistakenly decide that the target had been presented in the set. They also showed poor memory for the sizes of the individual circles in the set. Ariely's interpretation was that the visual system automatically constructs an average representation of a set of similar stimuli, but that information about the individual exemplars is lost.

Set averaging has been demonstrated for many low-level perceptual properties, including size (e.g. Ariely, 2001; Chong & Triesman, 2003, 2005; Albrecht & Scholl, 2010; Corbett and Oriet, 2011; Marchant, Simons & de Fockert, 2013), orientation (e.g. Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) and spatial location (Alvarez & Oliva, 2008). It also seems to occur with more complex stimuli, such as faces.

Early research on this topic within the domain of face processing investigated "prototype abstraction". Solso and McCarthy (1981) asked participants to memorise Identikit faces that diverged from unseen prototypes in terms of one to four "features" (hair, eyes, nose and chin, or mouth). In an old/new recognition test that included the prototypes, the prototypes were falsely recognised as having been seen before - and participants were more confident in these decisions than in their judgements about some of the faces that had actually been presented before.

Bruce, Doyle, Dench and Burton (1991) used artificial faces generated with a computerised face construction system called "Mac-a-Mug Pro". Whereas Solso and McCarthy varied facial features, Bruce et al manipulated the facial configuration: variants on a face differed only in terms of the vertical location of their facial features within the face shape. In an incidental learning paradigm, participants were exposed to the variants (but not the prototype from which they were derived). When presented with a distractor and the prototype, participants showed a strong preference for deciding that the prototype had been seen before. This was true only when the distractor was a variant that differed markedly from the variants presented in the learning phase; if the distractor was similar to the prototype, participants were unable to decide which had been shown previously.

Cabeza, Bruce, Kato and Oda (1999) conducted a series of experiments using photographic-quality face images, and morphs (blends) between individual faces. They were primarily interested in how variations in similarity between face images affected participants' ability to detect whether the images had been shown before (image recognition) and whether they showed the same person (identity recognition). Prototype effects were again observed. Anticipating the present research, Cabeza et al. noted that even when morphed exemplars were so dissimilar from each other that they were effectively different individuals, participants still showed a strong tendency to falsely recognise prototypes as faces that had been seen before.

Haberman and Whitney (2007, 2009, 2012a,b) showed that participants were able to accurately extract the mean emotion from a set of expressive faces, without being able to identify any of the individual faces. De Fockert (de Fockert & Wolfenstein 2009; de Fockert & Gautrey 2013) demonstrated perceptual averaging for facial identity: after viewing sets of four faces, participants tended to made a greater proportion of “present” (i.e. previously seen) responses for morphs of the sets than they did for genuine set members. Again, the interpretation is that averaging occurred at the expense of a loss of information about individual exemplars.

While these studies demonstrate fairly conclusively that some kind of averaging of facial information occurs, two issues warrant further investigation. Firstly, is averaging really accompanied by a loss of information about the individual set members? Since most of the earlier studies used highly similar stimuli, it is perhaps unsurprising that participants failed to remember the individual exemplars (Neumann, Schweinberger & Burton, 2013).

Secondly, in the case of faces, what exactly is being "averaged"? A limitation with the earlier studies is that they used highly similar images at presentation and test. Consequently de Fockert and Wolfenstein (2009) could not exclude the possibility that averaging might be based on low-level information, such as texture. There is already some evidence that face averaging involves more than this. Although their images were variations on a single face, Haberman and Whitney (2009) showed that set averaging was significantly more likely to occur with upright faces than with inverted or scrambled faces. This difference is hard to explain if averaging was based solely on low-level information. Neumann, Schweinberger and Burton (2013) used de Fockert and Wolfenstein's (2009) procedure, but with famous (celebrity) faces. On trials in which the target face was either a set member or a morph of the set, the target was either identical to the set in terms of viewpoint (image matching), or in a different viewpoint (identity matching). Participants still incorrectly recognised the morphs as set members, even when they were presented in a different viewpoint.

The studies by Haberman and Whitney (2009) and Neumann et al. (2013) demonstrate that the extraction of mean identity is not necessarily based on low-level visual information. However, familiar and unfamiliar face recognition are known to differ in various ways, the former being based more on abstractive representations and the latter on more "image-based" representations (Megreya and Burton 2006). To our knowledge, only two other studies have investigated set averaging of identity with different views of unfamiliar faces. Rhodes et al (2014) examined set averaging in children, in a study whose primary aim was to compare the process in normally developing children and children with autism. Using different images at presentation and test reduced the size of the set averaging effect compared to when identical images were used. Leib, Fischer, Liu, Qiu, Robertson and Whitney (2014) demonstrated viewpoint-invariance in set averaging with unfamiliar faces, but using a rather different technique from the usual method. Participants saw a set of different faces (based on morphs between three familiar faces) and then attempted to select the mean of this set of faces from a large array of 144 faces. Since the set faces were all morphs of three familiar faces, and the task was to *explicitly* identify the mean face on each trial, further research is warranted to support Leib et al's claim that set averaging occurs with unfamiliar faces despite viewpoint changes.

The present experiment examined perceptual averaging using de Fockert's technique with unfamiliar faces that had been vertically stretched to three times their normal height (see fig. 1). This manipulation has remarkably little effect on face recognition, despite its marked effects on the configural information that is widely considered to underlie recognition. Stretching does not impair explicit recognition (using famous/nonfamous decisions: Hole, George, Eaves & Rasek 2002); stretched faces do not produce dishabituation as measured by ERP activity (Bindemann, Burton, Leuthold & Schweinberger, 2008) or fMRI activity in face-selective brain regions (Baseler, Young, Jenkins, Burton & Andrews 2016); and face identity after-effects for stretched and normal faces are identical (Hole 2011). Using stretched faces within the set averaging paradigm offers an additional method for comparing the processing of normally-proportioned and stretched faces, to explore whether stretched and normal faces are processed in comparable ways in every respect. On the basis of the previous research on stretching, we would predict that set averaging will occur despite stretching of either the target face or the set members.

More importantly, using stretched faces provides a means to divorce low-level physical information from high-level identity information. Vertically stretching faces markedly changes their low-level physical properties. If set averaging occurs despite stretching (either of the set members or the target face) then this implies that averaging is occurring at the level of high-level identity processing rather than low-level image processing.

Finally, previous studies have measured participant's memory for set members and averages only indirectly, by measuring the speed and accuracy to decide whether a face had been seen before. As an additional test of whether responses to set members and set averages differed in any way, we also obtained ratings of participants' confidence in their decisions.

Method

Participants

Sixty Caucasian participants (25 male) were tested, aged between 18 and 35 years (*M* = 21 *SD* = 3). The study was approved by the University of Sussex Ethics Committee (application ER/AB595/1).

Design

An independent-measures design was used. Each participant was randomly allocated to one of three different conditions (N = 20 per condition):

- *Control* condition: normal (unstretched) face set followed by a normal (unstretched) target face ;

- *Stretched Set* condition: stretched face set followed by an unstretched target;
- *Stretched Target* condition: unstretched face set followed by a stretched target.

The main dependent variable was the number of times each participant responded "yes" when asked to determine whether or not a target face was present in a preceding set. Their confidence in each of their responses was also recorded using a 7-point Likert scale.

Apparatus and Stimuli

Stimuli were based on 90 greyscale photographs of male Caucasian faces in frontal views, with no facial hair or piercings and with a neutral expression. These were taken from the Glasgow Unfamiliar Face Database (Burton, White & McNeill, 2010) and the Aberdeen database (*pics.stir.ac.uk*). Using Adobe Photoshop Elements 6.0, each face was cropped to its internal features by applying an oval mask. Forty of the faces were used to produce ten sets of four faces. Four different types of face target were then created for each set:

Matching single face: one face from each face set was randomly selected to be a target.

Matching morph: for each face set, a matching morph was generated with SmartMorph Version 1.55 (Vinther, 2004) using a similar method to de Fockert and Wolfenstein (2009). For each set, two morphs were firstly created, each combining two faces (e.g. A with B, and C with D). To morph two faces together, 80 markers were placed around the eyes, eyebrows, nose, mouth and the outline of each face. The positions of markers on one face were individually aligned to corresponding locations on the other face. The AB and CD morphs were then morphed together, to produce an average face based on all the four faces of a set. Using this method, 10 male morphed faces were created. Finally, a frame was placed around each face.

Non-matching single faces: for each face set, a face was selected from the unused pool of 50 faces and randomly assigned to that set.

Non-matching morphs: from the remaining unused 40 faces, 10 additional sets of four faces were created. These sets were not seen by the participants. The same method used to create the matching morphs was applied, to produce 10 morphed faces. Each of these was randomly assigned to one of the sets of faces that would be seen by the participant.

Each target face was surrounded by a rectangular black frame, to differentiate it from the preceding face set when it was displayed.

To produce the stretched face conditions, each of the 10 sets of faces, including their 4 target faces (i.e. matching morph, non-matching morph, matching face and non-matching single face) were vertically stretched to three times their original height, while preserving their original width.

Pilot study

The morphing procedure produces faces with slightly less definition than real faces. To eliminate this as a cue, the target matching and non-matching single faces were blurred slightly, using the 'Gaussian blurring tool' in Adobe Photoshop Elements 6.0.[1] A pilot study confirmed that the target morphs could not be distinguished from the target real faces purely on the basis of their low-level image properties. Twelve participants each viewed 80 faces (20 stretched morphs, 20 stretched single faces, 20 non-stretched morphs and 20 non-stretched

[1] The blurring tool was set to a value of 1.5 pixels for stretched faces and 2 pixels for non-stretched faces. Less blur was needed for the stretched faces because stretching in itself reduces image resolution slightly.

single faces) that were individually flashed at the centre of the screen for 500 ms. After each face they were asked "*Was that preceding face a MORPH or a SINGLE face ?*" Overall, performance was significantly below chance level (50% correct decisions): *M* = 41%, *SD* = 8.30, $t(11) = -3.87$, $p = .003$, $d = 1.12$. Inspection of the pattern of judgements for each face-type suggests that there was a bias towards perceiving most of the stimuli as "real" faces rather than as morphs. On average, participants misclassified 62% of the stretched morphs and 77% of the unstretched morphs as being single faces. 66% of the unstretched single faces were correctly identified as such, but only 36% of the stretched single faces (i.e., 64% of them were wrongly thought to be morphs). Perhaps in the case of the stretched single faces, the slight softening of definition produced by stretching was misinterpreted as evidence that these faces were morphs.

Procedure

Set averaging task. The experiment was run on a Microsoft Surface Pro 10.6" Tablet running E-Prime 2.0 Professional (Psychology Software Tools, Inc., 2012). There was one session of forty trials. Each of the 10 sets of faces appeared four times, followed on each occasion by a different target face. The experiment started with an instruction screen and four practice trials. Trials appeared in a different random order for each participant. On each trial, four faces appeared sequentially at the centre of the screen for a duration of 500ms each. This was followed by a fixation cross (250ms) and a framed target face (500ms) (see figure 1). Similarly to de Fockert and Wolfenstein's (2009) procedure, the target face was either:

(i) a matching morph (a morph of the preceding four faces);

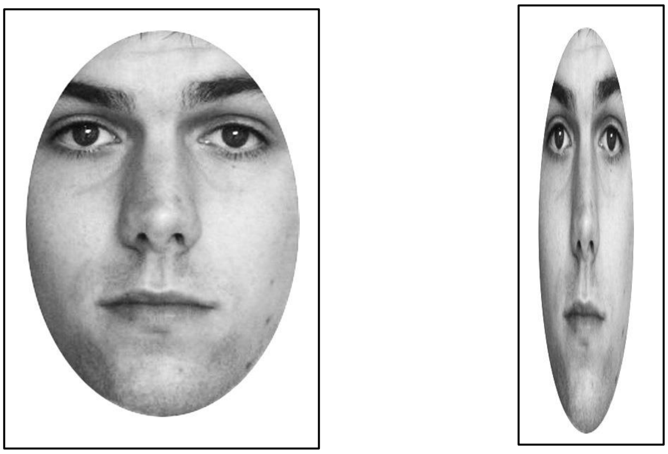
(ii) a non-matching morph (a morph of a different set of faces);

(iii) a matching single face (one of the preceding four faces); or

(iv) a non-matching single face (a face that was not one of the preceding four faces).

After the target face, the question "*Was that face present in the preceding set?*" appeared on the screen and remained until the participant made a response ("Y" for Yes and "N" for No). Once participants responded, the question "*On a scale of 1 (not at all confident) to 7 (very confident) how confident are you with your response?*" appeared on the screen and remained until the participant indicated their confidence level, using the numbers 1 to 7 on the keyboard. After answering that question, the next trial started.

(a)



(b)

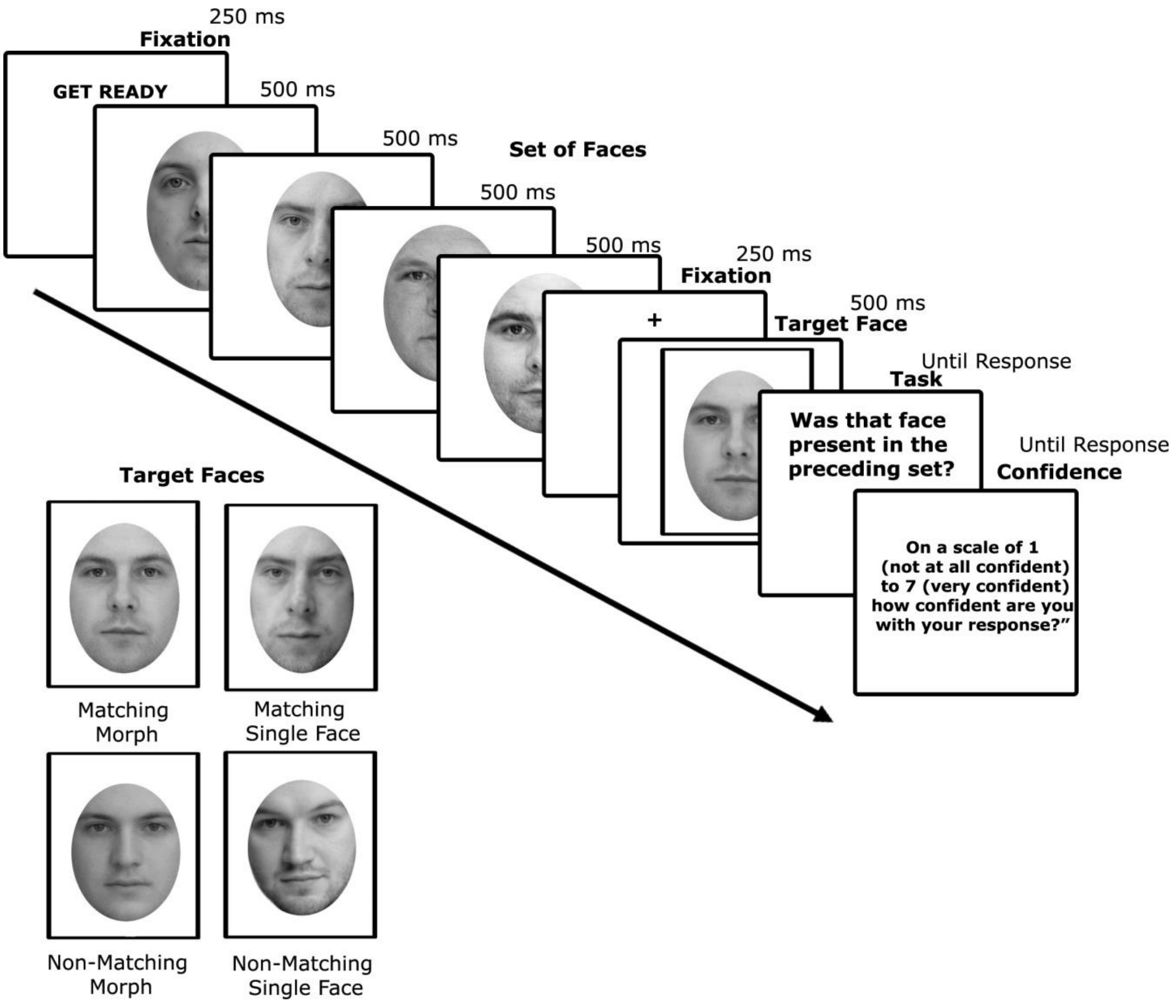


Figure 1. (a) A set face in normal and stretched versions. (b) A sample trial consisting of a set of four face identities and a target face. Target faces could be either a single face or a morph, that either matched or did not match the preceding set of four faces.

Results

Percentage of "Yes" Responses.

From the total number of "yes" responses that each participant made, the percentage of "yes" responses attributed to each target face (i.e. matching member, matching morph, non-matching morph, and non-matching member) was calculated. This was done for all three conditions. Because these percentages were not normally distributed, the data were log transformed and all data analyses were conducted using the transformed data. Following Bland and Altman's (1996) recommendations, (see also Hale, 2014; McDonald, 2014) means and confidence intervals were back transformed, but not standard errors.

Fig. 2 shows the mean percentage of "yes" responses for each type of target face, for all three conditions. A 2 (Face Transformation: morph vs. single) x 2 (Face Membership: matching vs. non-matching) x 3 (Condition: Control vs. Stretched Set vs. Stretched Target) mixed-measures ANOVA was conducted on the percentage of "yes" responses, with Condition as the between-subjects factor.

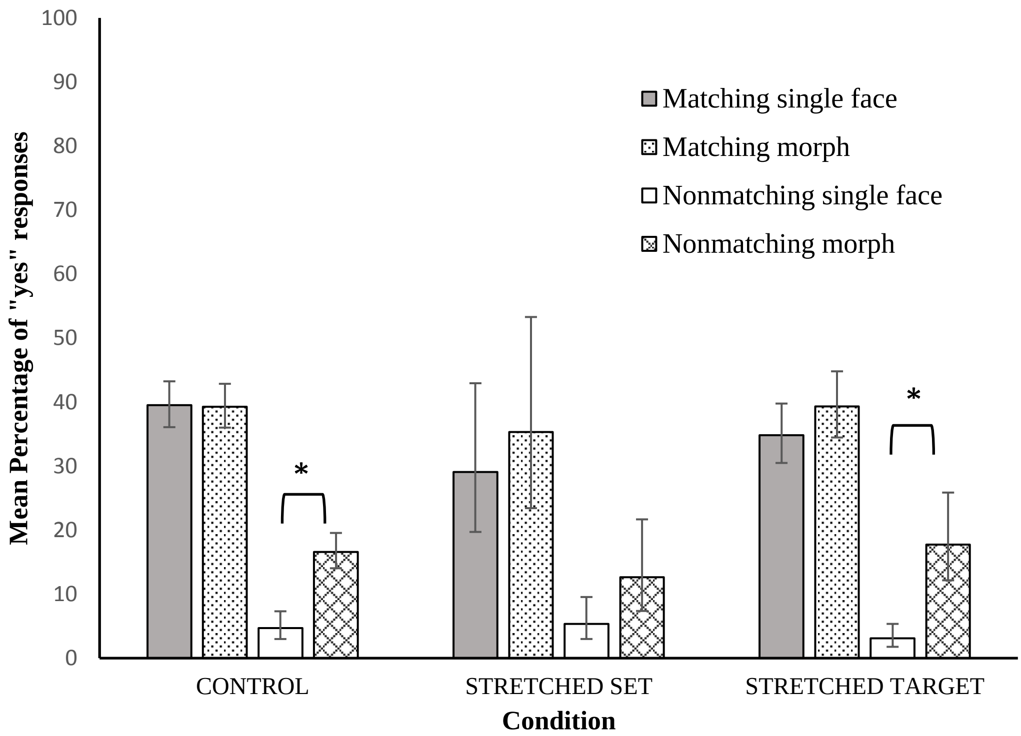


Figure 2. Mean percentage of "Yes" responses for each type of face target in each condition (error bars represent 95% Confidence Intervals: * $p < .001$).

Stretching faces had no effect on the pattern of percentage of "yes" responses. There was no significant effect of Condition, either as a main effect or in interaction with Face Transformation or Face membership, all *F*s <= 1.

The effect of Face Membership was significant, $F(1,57) = 150$, $p < .001$, $\eta_p^2 = .73$. Participants made significantly more "yes" responses to matching target faces ($M = 36\%$, 95% CI [33, 39]) than to non-matching target faces (*M* = 8%, 95% CI [7, 10]). This indicates that participants could distinguish faces that belonged to previously-seen face sets from faces that belonged to unseen face sets. There was also a main effect for Face Transformation, $F(1, 57) = 39.51$, $p < .001$, $\eta_p^2 = .41$. Overall, participants made significantly more "yes" responses to morphed faces (*M* = 24%, 95% CI [22, 27]) than to single faces (*M* = 12%, 95% CI [10 , 14]). At first sight this appears to imply that average faces seemed more familiar to participants than single faces. However a significant interaction between Face Transformation and Face Membership, $F(1, 57) = 44.30$, $p < .001$, $\eta_p^2 = .44$, suggests that this was true only for non-matching face targets. Non-matching morphs received more "yes" responses than non-matching single faces, $t(59) = 7.36$, $p < .001$[2], $d = 0.95$, but there were no corresponding differences between matching morphs and matching single faces, $t(59) = 1.00$, *n.s.*. Participants were as likely to decide (mistakenly) that a matching morph had been seen in the preceding set as they were to decide (correctly) that a matching single face had been seen previously.

Confidence Ratings.

Not every participant made “yes” and “no” responses for each possible permutation of stimulus set and target face. For example, when presented with a non-matching single target face, some participants never responded that this face had been presented in the preceding set, and thus produced no confidence data for that response category. If a participant had missing data, these were replaced with the mean for that response category, calculated from the data of the other participants in the same condition (*control*, *stretched set* or *stretched target*) who did provide data.

[2] All pairwise comparisons in this paper are Bonferrroni-corrected.

We performed two 2 (Face Transformation: morph vs. single face) x 2 (Face Membership: matching vs. non-matching) x 3 (Condition: *control* vs. *stretched set* vs. *stretched target*) mixed-measures ANOVAs on the confidence data. One ANOVA was conducted on the confidence for "yes" decisions (deciding that a face was a set member) and the other ANOVA was conducted on the confidence for "no" decisions (deciding that a face was not a set member).

For the "yes" responses, there were significant effects of Face Membership, $F(1, 57) = 25.04$, $p < .001$, $\eta_p^2 = 0.31$, and Condition, $F(2, 57) = 4.99$, $p = .01$, $\eta_p^2 = 0.15$. There was also a significant interaction between Face Membership and Face Transformation, $F(1, 57) = 5.34$, $p = .02$, $\eta_p^2 = .09$. No other effects were significant, all *F*s < 1.9. Inspection of fig. 3a suggests that participants in all three conditions showed a broadly similar pattern. They were most confident in responding "yes" to faces that were matching faces, slightly less confident with responding "yes to matching morphs, and least confident in responding "yes" to non-matching single faces or morphs. Tukey HSD tests revealed that overall confidence in the *control* condition ($M = 5.07$, 95% CI [4.7, 5.4]) was significantly higher than in the *stretched set* condition, $p = .008$, but not the *stretched target* condition, p = .58. The two stretched conditions did not differ from one another, $p = .10$ (*stretched set*: M = 4.27, 95% CI [3.9, 4.6]; *stretched target:* M = 4.81, 95% CI [4.4, 5.2]). Note however that in absolute terms, confidence was reasonably high in all three conditions.

Fig. 3(a) appears to show that participants in the *control* condition were more confident in their "yes" responses to matching faces (M =5.63, $SD = 0.67$) than to matching morphs ($M = 5.23$, $SD = 1.02$). However, this difference failed to reach significance with a Bonferroni-corrected paired samples *t*-test, $t(19) = 2.15$, p =.04. This was also true for the *stretched target* condition (matching faces: $M = 5.15$, SD = 0.72; matching morphs: $M = 4.77$, *SD* = 0.80; *t*(19) = 2.35, *p* = .03) and the *stretched set* condition (matching faces: *M* = 4.62, *SD* = 1.19; matching morphs: *M* = 4.39, *SD* = 1.18; *t*(19) = 0.95, *n.s.*).

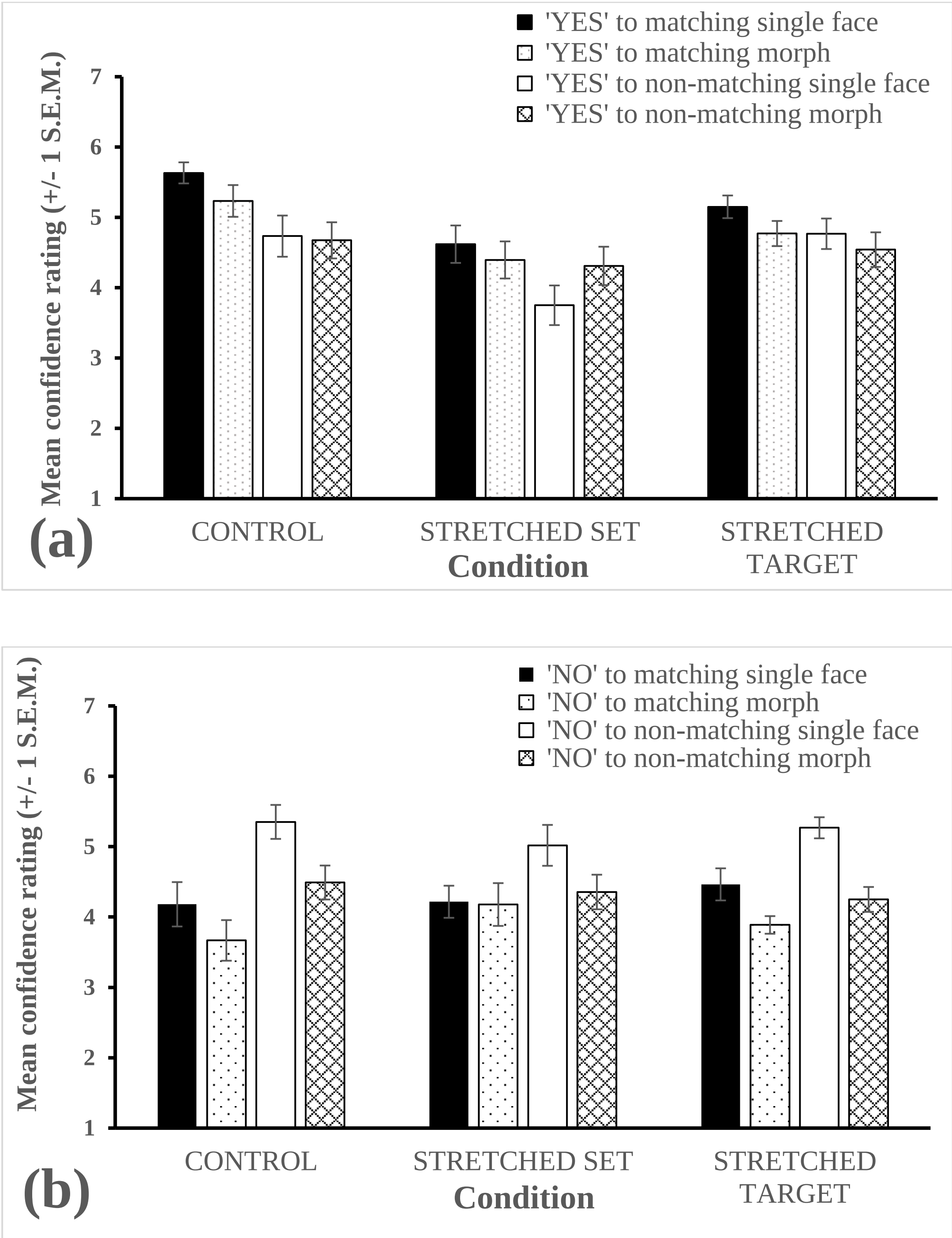


Figure 3. Mean confidence ratings in each condition (from 1= *not at all confident* to 7 = *highly confident*) for (a) "Yes" responses (deciding that a face had been displayed in the preceding face set) and (b) "No" responses (deciding that a face had not been displayed in the preceding face set).

For the "no" responses, there were significant main effects of Face Membership, $F(1,57) = 71.09$, $p < .001$, $\eta_p^2 = 0.56$ and Face Transformation, $F(1, 57) = 38.98$, $p < .001$, $\eta_p^2 = 0.41$. There were also significant interactions between Face Membership and Face Transformation, $F(1, 57) = 4.97$, $p = .03$, $\eta_p^2 = 0.08$ and between Face Membership and Condition, $F(2, 57) = 3.62$, $p = .03$, $\eta_p^2 = 0.11$. No other effects were significant, all *F*s <1.87. Inspection of fig. 3b shows that, in all three conditions, participants were most confident in saying "no" to faces that were non-matching faces. "No" responses to matching faces, matching morphs and non-matching morphs were made with lower confidence, although again, in absolute terms, participants were reasonably confident in all their decisions.

Discussion

This experiment examined the effects of vertical stretching on the extraction of mean identify from a set of unfamiliar faces. Three main findings were noted. Firstly, we replicated the set averaging effect: when shown a "set morph" that represented the mean identity of a set of faces, participants tended to mistakenly report that this had been presented in the set. Secondly, set averaging was unaffected by either vertical stretching of the set faces or the target. Thirdly, participants were no more confident in deciding that a target face had been presented in the preceding face set when it was a set member than when it was a set morph.

Across all conditions, the percentages of “yes” responses for set morphs were the same as for set members. "Yes" responses for non-matching faces and non-matching morphs were given significantly less frequently. This pattern of results is consistent with previous experiments (e.g. de Fockert & Wolfenstein, 2009) suggesting that face identities may be summarised to create a stable face representation. However, unlike de Fockert and Wolfenstein (2009) and de Fockert and Gautrey (2013), in the present experiment, matching morphs did not receive significantly more “yes” responses than genuine set members. This difference could be explained by the stimuli used: de Fockert and Wolfenstein (2009) created sets of faces that were similar in terms of features and skin tone. Similarities between stimuli in a set can increase the likelihood of recognition errors (Busey & Tunnicliff 1999; Cabeza et al, 1999; Neumann et al., 2013). In the current experiment, the sets of faces were more variable. Consequently, in our experiment, as well as forming a single summary representation, observers also had a good memory for single faces. Our findings extend those of Neumann et al. (2013), who found that the extraction of mean identity from familiar faces was possible despite participants having a good memory for the set members. These results question the argument that summary representations help to overcome poor memory for single items in a scene (Neumann et al., 2013; Oliva & Torralba, 2006).

Wallis, Siebeck, Swann, Blanz and Bülthoff (2008) suggest that it is useful to distinguish between two different versions of the "prototype effect". A "Level 1" effect occurs if participants consider the prototype face to be more familiar than other novel faces. A "Level 2" effect occurs if the prototype face is additionally judged to be more familiar than the faces which have been seen before. Using a procedure broadly similar to Solso and McCarthy's (1981) study, but with computer-generated face models that varied in the number of different "face regions" rather than isolated "features", Wallis et al. found evidence for a Level 2 effect. In our study, using the set averaging process with real individual faces, we found no evidence for anything beyond a Level 1 effect: participants were likely to decide that a matching morph had been seen before, but their overt decisions and their confidence in those decisions (see below) provided no evidence that they were *more* likely to do this for matching morphs than for faces that they had genuinely encountered before.

Stretching did not affect participants' ability to correctly determine whether or not a target face was present in the preceding set. This is consistent with previous research demonstrating that geometric distortions, such as stretching, do not impair recognition accuracy (Hole et al., 2002; Sandford et al., 2013). These previous studies used famous/non-famous judgment tasks. The current experiment extends these previous findings by demonstrating that stretching also had no effect on the recognition of unfamiliar faces during a face-matching task. Because stretching changes the original aspect ratio and absolute distances between facial features, the obtained findings further question the importance of inter-attribute distances for face identification (Maurer et al., 2002; Sandford & Burton, 2014; review in Burton et al., 2015).

In the *stretched set* and *stretched target* conditions, participants extracted summary statistics from identity despite the gross changes in image properties induced by stretching. These findings further support the idea that ensemble coding of faces can occur at the level of relatively high-level identity information, rather than merely involving low-level image properties (Leib et al, 2014; Neumann et al., 2013). Importantly, the current experiment investigated ensemble coding implicitly and not explicitly as in Leib et al.'s (2014) study, which suggests that observers can use distorted information to formulate a mean representation automatically and not only when instructed to do so.

In analysing the confidence data, we were interested in two questions. Firstly, when participants (mistakenly) identified the set morph as being one of the faces that they had seen in the preceding set, were they as confident in this decision as when they correctly identified a set face as being one of the faces in the set? It appears that they were. Confidence ratings for "yes" responses for actual set members were relatively high, and not significantly different from those for set morphs, in any of the three conditions (*control, stretched target* or *stretched set*). The confidence data thus corroborate the impression obtained from the "yes/no" response data, that the set averages (morphs) are indistinguishable from individual set members.

The confidence ratings reported in this experiment are consistent with previous research investigating recognition errors. Solso and McCarthy (1981) found that participants reported high positive confidence ratings for prototype faces in comparison to previously studied faces. Furthermore, after studying a series of individual faces, participants were more confident in their “previously studied” than “not studied” responses when presented with conjunction faces (novel faces representing the combination of features from studied faces) (Reinitz et al., 1992). However, during these experiments, faces were studied for 10-30 seconds each, whereas the present experiment displayed faces for half a second each. This suggests that participants can still be confident in their responses even when not meticulously learning the stimuli (Haberman & Whitney, 2007).

A second question is whether stretching (either of the set or of the target face) affected the confidence with which participants decided that a single face or a set morph was a set member. Again, differences between the conditions were small in absolute terms, although response confidence was significantly lower in the *stretched set* condition than in the *control* and *stretched target* conditions. The small size of these differences suggests that stretching faces did not encourage participants to guess more than when they were presented with normally-proportioned faces.

Hitherto, the set averaging phenomenon has been investigated primarily because of interest in the averaging phenomenon itself, and speculation about whether or it has any functional role in recognition. However it can also be regarded as a potentially useful technique for examining the informational underpinnings of face recognition. By investigating the conditions under which the averaging process does or does not take place, it may be possible to obtain information about the very basis of face recognition. Vertical stretching did not affect the set averaging process, consistent with previous research showing non-effects of stretching on explicit recognition (e.g. Hole et al., 2002). However face recognition is impaired by other transformations, such as non-global face distortions (Hole et al, 2002), photographic negation (Galper, 1970) and inversion (Yin, 1970), and also by changes in pose and expression. Future research could investigate how these transformations affect the set averaging process. Comparing the effects of these transformations would help demonstrate how different forms of facial information might be important in extracting mean identity (Taschereau-Dumouchel et al., 2010). Additionally, this would also help demonstrate whether or not face recognition and the extraction of mean identity are impaired by similar distortions/transformations, and test Leib et al's (2012) suggestion that face recognition and the extraction of summary statistics are independent processes[3]. The present study used only frontal views of faces. To further support the claim that participants were extracting the mean *identity*, rather than the mean of low-level featural information, faces photographed from different angles could be used to create an identity-matching task (Neumann et al., 2013).

In conclusion, observers were able to unite different identities over time to form a stable face representation. These summary representations were recognized just as often as genuine faces, and more importantly, tolerated an affine linear transformation (stretching). These results add to the evidence that summary representations can be based on relatively high-level perceptual information and they suggest that facial recognition might not exclusively rely on configural information. This technique may be a useful tool for investigating how efficiently different facial characteristics can be encoded by the visual system.

Acknowledgements

[3] Face recognition ability did not correlate with the ability to discriminate mean face identity (Leib et al., 2012).

We are very grateful to Peter Hancock and de Fockert for their thorough and insightful reviews of earlier drafts of this article.

References

Albrecht, A.R. & Scholl, B.J. (2010). Perceptually averaging in a continuous visual world: extracting summary representations over time. *Psychological Science*, *21*(4), 560-567.

Alvarez, G. A. & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392-398.

Ariely, D. (2001). Seeing sets: representation by statistical properties. Psychological Science, 12(2), 157–162.

Baseler, H.A., Young, A.W., Jenkins, R., Burton, A.M.. & Timothy J. Andrews, T.J. (2016). Face-selective regions show invariance to linear, but not to non-linear, changes in facial images. *Neuropsychologia, 93*, 76-84.

Bindemann, M., Burton, A.M., Leuthold, H., & Schweinberger, S.R.. (2008). Brain potential correlates of face recognition: geometric distortions and the N250r brain response to stimulus repetitions. *Psychophysiology 45*, 535-544.

Bland, J. M., & Altman, D. G. (1996). Transformations, means, and confidence intervals. *BMJ: British Medical Journal*, *312*(7038), 1079.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition*, *38*, 109-144.

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M. (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), 482–496.

Burton, A. M., White, D. & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods*, *42(*1), 286-291.

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1210- 1235.

Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: extension and limits. *Memory & Cognition*, *27*(1), 139-151.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393-404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: computing the average size in perceptual groups. *Vision Research*, *45*(7), 891-900.

Corbett, J.E. & Oriet, C. (2011). The whole is indeed more than the sum of its parts: perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*, 289-301.

de Fockert, J. W., & Gautrey, B. (2013). Greater visual averaging of face identity for own-gender faces. *Psychonomic Bulletin & Review*, *20*(3), 468-473.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1716-1722.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science*, *19*(4), 207-208.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*, 751–753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718-734.

Haberman, J. & Whitney, D. (2012). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception and Psychophysics*, *72*(7), 1825-1838.

Haberman, J. & Whitney, D. (2012). "Ensemble Perception: Summarizing the scene and broadening the limits of visual processing." In J. Wolfe and L. Robertson (Eds.), *Perception to Consciousness: Searching with Anne Treisman.*(pp.339-349) Oxford University Press.

Haig, N. D. (1984). The effect of feature displacement on face recognition. *Perception*, *13*(5), 505-512.

Hale, I. (2014). Topic 8: Data Transformation. Retrieved February 2, 2015, from the University of New Hampshire Website: http://www.unh.edu/halelab/BIOL933/lectures/lect_11_reading.pdf

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, *31*(10), 1221-1240.

Leib, A. Y., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: a viewpoint invariant mechanism to represent average crowd identity. *Journal of Vision*, *14*(8), 1-13.

Leib, A. Y., Puri, A. M., Fischer, J., Bentin, S., Whitney, D., & Robertson, L. (2012). Crowd perception in prosopagnosia. *Neuropsychologia*, *50*(7), 1698-1707.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, *6*(6), 255-260.

McDonald, J.H. (2014). *Handbook of Biological Statistics* (3rd ed.). Baltimore, MD: Sparky House Publishing.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: evidence from a matching task. *Memory & Cognition*, *34*(4), 865-876.

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*(1), 56-63.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in Brain Research*, *155*, 23-36.

Psychology Software Tools, Inc. (2012). E-Prime Professional (Version 2.0) [Computer Software]. Sharpsburg, PA : Psychology Software Tools, Inc

Reinitz, M. T., Lammers, W. J., & Cochran, B. P. (1992). Memory-conjunction errors: miscombination of stored stimulus features can produce illusions of memory. *Memory & Cognition*, *20*(1), 1-11.

Rhodes, G., Neumann, M.F., Ewing, L. & Palermo, R. (2015) Reduced set averaging of face identity in children and adolescents with autism. *Quarterly Journal of Experimental Psychology, 68*(7), 1391-1403.

Sandford, A., & Burton, A. M. (2014). Tolerance for distorted faces: challenges to a configural processing account of familiar face recognition. *Cognition*, *132*(3), 262-268.

Sandford,A., Burton, A. M., & White, D. (2013) Transparency of geometric distortions in face identity. *i-Perception, 7*(4), 480.

Solso, R. L., & McCarthy, J. E. (1981). Prototype formation of faces: A case of pseudo-memory. *British Journal of Psychology*, *72*(4), 499-503.

Taschereau-Dumouchel, V., Rossion, B., Schyns, P. G., & Gosselin, F. (2010). Interattribute distances do not represent the identity of real world faces. *Frontiers in Perception Science*, *1*, 1-10.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107-141.

Vinther, M. (2004). Smartmorph (Version 1.55) [Computer Software]. Enfield, UK :MeeSoft.

Wallis, G., Siebeck, U. E., Swann, K., Blanz, V., & Bülthoff, H.H. (2008). The prototype effect revisited: evidence for an abstract feature model of face recognition. *Journal of Vision*, *8*(3): 20, 1-15.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141-154.