Image Dependency In The Recognition Of Newly Learnt Faces

Christopher A. Longmore
University of Plymouth, UK
University of Exeter, UK

Isabel M. Santos
Carlos F. Silva
University of Aveiro, Portugal

Abi Hall
Dipo Faloyin
Emily Little
University of Exeter, UK

Address of correspondence:

Chris Longmore
School of Psychology
Faculty of Health and Human Sciences
Plymouth University
Drake Circus
Plymouth
PL4 8AA

Tel: +44 (0)1752 584890
Fax: +44 (0)1752 584808

Email: chris.longmore@plymouth.ac.uk

Abstract

Research investigating the effect of lighting and viewpoint changes on unfamiliar and newly learnt faces has revealed that such recognition is highly image dependent and that changes in either of these leads to poor recognition accuracy. Three experiments are reported to extend these findings by examining the effect of apparent age on the recognition of newly learnt faces. Experiment 1 investigated the ability to generalise to novel ages of a face after learning a single image. It was found that recognition was best for the learnt image with performance falling the greater the dissimilarity between the study and test images. Experiments 2 and 3 examined whether learning two images aids subsequent recognition of a novel image. The results indicated that interpolation between two studied images (Experiment 2) provided some additional benefit over learning a single view, but that this did not extend to extrapolation (Experiment 3). The results from all studies suggest that recognition was driven primarily by pictorial codes and that the recognition of faces learnt from a limited number of sources operates on stored images of faces as opposed to more abstract, structural, representations.

Keywords: face recognition, pictorial coding, structural coding, image-dependency

Introduction

Familiar face recognition has been demonstrated to be highly robust whilst the recognition of unfamiliar faces is poor (Johnston & Edmonds, 2009). The key difference between the two forms of recognition is that unfamiliar face recognition appears to be dependent upon the similarity between studied and subsequent test images whilst familiar face recognition shows no such bounds. For example, Bruce (1982) presented participants with images of familiar faces and in a later test, participants were able to accurately indicate which *people* had been previously seen despite the *image* itself changing in pose and expression. The lack of image dependency for familiar faces has also been demonstrated by Hole, George, Eaves and Rasek (2002) who presented participants with images of familiar faces that had been vertically stretched to twice the normal height and found that recognition accuracy was not affected by this manipulation. Another example of the robustness of familiar face recognition is peoples' ability to recognise individuals across naturalistic ageing (Bruck, Cavanagh, & Ceci, 1991). Bruck et al. presented participants with photographs of their former fellow high school students that were photographed 24 to 26 years prior to testing and asked them to match them to a current photograph of the individual. The results demonstrated that participants who were familiar with the individuals were able to perform this task indicating that familiarity with a face aids recognition across different ages despite the large changes in the pictorial representation of an individual after around 25 years. These results point towards a system for familiar face recognition that is not dependent upon any particular seen image and it has been suggested that a structural code, created via the means of an abstract structural representation of the face, similar to that proposed for object recognition by Marr and Nishihara (1978), is constructed over

repeated exposures to a particular face and used for subsequent recognition. It has been proposed as a representation of the face that is consistent across transformations by extracting image-invariant elements of the face (Bruce, 1982). This is assumed to enable the robust levels of recognition associated with familiar face recognition (Bruce & Young, 1986; Jenkins, White, Van Montfort, & Burton, 2011).

In contrast to the robustness of familiar face recognition, unfamiliar face recognition is fragile. Bruce (1982) presented participants with images of unfamiliar faces and found that in a later recognition test, images that were identical to the studied image were recognised well, but changing the pose and/or the expression between study and test had a considerably detrimental effect on recognition accuracy. Bruce suggested that unfamiliar face recognition might be poor due to a reliance on pictorial codes (codes that are specific to a studied image or instance of a face) on the basis that it appears that changing the image of an unfamiliar face from the studied instance, whether in an experimental situation (e.g. Bruce, 1982; Hill, Schyns, & Akamatsu, 1997; Krouse, 1981; Patterson & Baddeley, 1977) or more naturalistic settings such as an eye-witness attempting to recognise someone they saw only once (see Wells, 1993 for a reivew) leads to poor recognition accuracy.

Studies investigating unfamiliar face memory have typically presented participants with a single image of an unknown face (e.g. Bruce, 1982). However even when an image of a face is learnt extensively recognition tends to be poor. Longmore, Liu, and Young (2008) presented images of unfamiliar faces multiple times and from multiple viewpoints. Their results consistently showed that whilst multiple exposures to a facial image increase overall accuracy against viewing an image once, the ability to generalise to

novel viewpoints was not aided by the repeated presentation of the studied image combined with the studying of different images of the same individual. They concluded that despite the increased number of exposures and views of an individual's face, recognition was still primarily driven by an image-dependent pictorial code and suggest that the face learning process operates via the accumulation of these pictorial codes.

An accumulation of pictorial codes might account for other reported findings. Results from research on person identification via matching tasks have regularly found that errors in unfamiliar face matching are high (Burton, White, & McNeill, 2010) even when conditions are optimal (Bindemann, Avetisyan, & Blackwell, 2010). For example, Kemp, Towell, and Pike (1997) investigated whether supermarket cashiers could accurately match a photograph present on a credit card to the card's user. The photograph could either depict the bearer of the credit card or someone else. Kemp et al. found that these professional cashiers accepted the "fraudulent" cards (i.e. credit cards depicting someone other than the user of the card) on more than 50% of occasions.

Part of the problem faced by participants in such experiments might be a limited number of images upon which to base their decision. To investigate this, Bindemann and Sandford (2011) asked participants to identify a target face, given in the form of a photo identification document, from an array of 30 images. After they had made their choice they were then given a second form of identification and, in conjunction with the first image, they were again asked to identify the target from the set of 30 images. Finally the process was repeated with a third image. With each additional image presented participants' performance increased suggesting that providing an increased

range of exemplars of a person's face, and thus an increased number of pictorial codes relating to an individual, can improve person identification.

The photo identification document used by Bindemann and Sandford (2011) displayed an image of the individual taken at most 19 months prior to the target image. This range in depicted age is small in comparison to the 10-year range that passport officers regularly have to make matches across (White, Dunn, Schmid, & Kemp, 2015) and these experts also make a substantial number of errors (White, Kemp, Jenkins, Matheson, & Burton, 2014) when matching photographs of unfamiliar people. Matching photographs across a 10-year range has provided challenging for observers. For example, White et al. (2015) report that recognition accuracy for images of faces that differed by an average of 9.7 years was low at around 45% correct. Despite the importance of the ageing process on a person's appearance, this transformation has received little attention to date. The experiments reported in this paper extend the investigation into the pictorial coding of newly learnt faces to examine the effect of age transformations. A range of images was created for each face by morphing together a picture of an individual when they were aged approximately 20, and another image from when they were aged around 60. From these source images, intermediate morphs were created producing *pseudo-ages* of 30, 40 and 50 years. The ability of the face processing system to generalise to novel images of an individual across an ageing transformation could then be explored.

Experiment 1

Experiment 1 investigated whether pictorial codes are the basis for the recognition of newly learnt faces that differ in age. In this experiment, participants were trained to

recognise a single image of a face using the same training procedure as Longmore et al. (2008, experiment 2). Their ability to recognise the same person from different images showing different ages was then tested. If recognition is reliant on pictorial codes, recognition performance across most (if not all) images should decrease with greater image changes from the studied picture. On the other hand, a structural code that contains image invariant information about the face would lead to more comparable levels of performance when generalisation is required across various changes in the image. It would still be expected that performance would still be best when recognising the originally studied image as both pictorial and structural codes could be used.

Method

Participants

Sixteen participants (5 males, 11 females) aged between 20 and 22 years (M = 21.063 years; SD = 0.772) took part in the study in return for course credit. All participants had normal or corrected to normal vision and none had any experience of the faces used in study.

Materials

Images of 20 individuals (10 males and 10 females) were used. For each individual, two images were obtained showing a full-face view of the person. One of the images depicted the individual when they were aged around 20 years and the other showed the person when they were aged approximately 60 years. None of the images depicted the individual with facial hair or wearing glasses. Each image was converted to greyscale and then edited to remove the background information to leave only the face visible, was resized to 384 pixels in height and the width of the image was expanded to 384 pixels whilst maintaining the aspect ratio. The background was then filled with a

homogenous grey. The pair of images for each individual were then morphed together using MorphX (Norrkross, 2009). Five combinations of the 20 and 60-year-old images of each individual were created resulting in a total of 100 images. The five combinations produced five pseudo-ages of 20, 30, 40, 50 and 60 years and were created by the following combinations of the 20-year-old and 60-year-old images respectively: 95% (20-year-old image)/5% (60-year-old-image), 77%/23%, 50%/50%, 23%/77%, and 5%/95%. These combinations were designed to yield a pseudo 20, 30, 40, 50 and 60-year-old image and the ratios were determined by a pilot study to reveal the optimal mix required for each age group across the full image set. An example of one individual is shown in Figure 1.

[Insert Figure 1 here]

The 20 identities were split into two sets containing 10 people each with an equal number of male and female identities in each set.

Apparatus

The images were displayed on a 17" LCD flat screen set to a resolution of 1280x1024 pixels and a display depth of 32-bits using a custom program written in Microsoft Visual Basic 6. Participants interacted with the program using a standard mouse.

Procedure

All participants completed the experiment in a quiet testing area and were sat approximately 60cms from the computer screen. For half the participants, set one of the faces acted as targets and the second set as distractors. For the other half of

participants, set two acted as targets and set one as distractors. All participants completed three phases: a presentation phase, a training phase and a test phase.

Presentation phase

In the presentation phase participants saw 10 faces depicting 10 different identities for 5 seconds each with 0.5 seconds between each face. Within this block of 10 faces, two images (one male, one female) depicted faces from a pseudo age of 20, two from a pseudo age of 30 and so on for each of the five pseudo ages. For each facial image, a name was presented beneath the face. The name/face pairings were randomly generated for each participant from a fixed set of 10 names (5 male and 5 female) with the only restriction being that the male names were only used for the male faces and the female names for female faces. Participants were instructed to study the face/name pairings during this stage.

Training phase

The training phase contained two blocks. In the first block the 10 faces were broken down into two sets of 5 faces, and in the second block all 10 faces were shown. The training task required participants to match a presented face to the correct name. A face was presented in the centre of the screen with possible names for the face presented on buttons located below the face. The participant clicked on the button that corresponded to the name they thought belonged to the face. Instant feedback was given to indicate whether the correct name had been selected or not. In cases where the participant made an incorrect choice, they were informed of the correct name. Any faces that were incorrectly identified were re-presented until all faces were correctly named. To move from the first set of 5 faces to the second, and from the second set of faces to the second

block of training, all 5 faces from each set had to be correctly named, without making a mistake, on three separate occasions. The second block of training followed the same procedure except all 10 faces were shown. The criterion for completing training was the same.

Test Phase

In the test phase participants were presented with 5 blocks of 20 faces, with each block comprising 10 images of targets and 10 distractors. Within each block, four faces (two targets and two distractors) were shown of each pseudo-age, ensuring that all pseudo-ages were shown an equal number of times. Across all 5 blocks, all targets and all distractors were shown at each of the 5 pseudo-ages. Participants were presented with a single face with two buttons marked *Yes* and *No*. Participants clicked *Yes* if they thought they recognised the person and *No* if they did not. Between each test block participants completed a top-up training phase in which all 10 originally studied target faces were presented and all faces needed to be correctly identified, with no errors, on three separate occasions. This top-up phase was undertaken to counter any effects of learning of the represented distractor identities.

Results

The mean proportion correct scores for each of the 5 learnt pseudo ages, across the 5 testing pseudo-ages, is shown in Figure 2.

[Insert Figure 2 here]

The mean proportion correct scores were entered into a two-way repeated measures ANOVA with learnt age and test age as independent factors. The Huynh-Feldt correction for departures from sphericity was used throughout the analyses and effect sizes were calculated using generalized η^2 (Bakeman, 2005). Effect sizes were measured against Cohen's recommendation of 0.02 for a small effect, 0.13 for a medium effect, and 0.26 for a large effect (Cohen, 1988 as cited in Bakeman, 2005). There were significant main effects of learnt age: $F(4.0,60.0) = 5.743$, $MSE = 0.069$, $\eta^2_G = .059$, $p < .001$ and of test age: $F(2.6,38.9) = 11.145$, $MSE = 0.073$, $\eta^2_G = .077$, $p < .001$, both with small-medium effect sizes. A significant interaction between learnt age and test age with a large effect size was also found: $F(9.9,148.2) = 52.449$, $MSE = 0.089$, $\eta^2_G = .645$, $p < .001$. Post-hoc analyses revealed that a change of 10 pseudo-years in the test view away from the learnt view yielded no significant change in recognition accuracy with the exception of learning 30 years and testing on 40 years ($p = .03$, Bonferroni corrections applied). In contrast, for a change of more than 10 pseudo-years either way (younger or older), all learning conditions yielded significantly lower performance (all p's < .05).

It is possible that the top-up phase in which participants were required to re-identify the target faces might have led to progressively improved performance across the five test blocks. To assess this, performance across the five test blocks was analysed. The mean proportion correct obtained in each of the five blocks is shown is Table 1.

[Insert Table 1 here]

The mean proportion correct for each of the five test blocks was analysed using a one-way repeated measures ANOVA. No effect of test block was found; F(4.0,60.0) = 0.536, MSE = 0.018, η_G^2 = .029, p = .710.

Figure 2 reveals that participants found certain learn/test combinations quite difficult. Chance performance in the current experiment was 50% (mean proportion correct = 0.5). Participant performance for each of these combinations was compared against this chance level using a one-sample t-test and the results are shown in Table 2.

[Insert Table 2 here]

Discussion

The results of Experiment 1 indicate that the recognition of faces learnt from a single photograph is highly image dependent, supporting the results of Longmore et al. (2008). The results also hint at the limits of the generalisation system for faces as for all learnt images some generalisation was possible, but this was limited to test images which contained no more than an approximately 20%-25% shift away from the learnt image and it may be that shifts greater than this are too different from the studied image to allow for effective generalisation. Participants found some learn view and test view combinations challenging, particularly for combinations that differed substantially in terms of pseudo-age. These difficulties resulted in such combinations yielding performance that was significantly below than what would be obtained by chance. This can be taken as evidence for the difficulty people have in generalising to different images of the same person (c.f. Jenkins et al., 2011). Indeed, participants only indicated

(correctly or incorrectly) that they had seen an individual previously on 30% of trials. It would appear that participants are not simply guessing whether an individual had been previously encountered or not, but rather were actively rejecting images of previous seen people due to limits of the face generalisation system.

Experiment 2

Experiment 1 revealed a strong image dependency in the recognition of faces learnt from a single image. However, learning a single image requires for all generalisation to be made from the studied picture as opposed to a more general representation of the face. Thus, Experiment 2 investigated whether learning two images of a face allows for better recognition on a novel image in comparison to learning only a single image. If such an effect is found, this would suggest that the face processing system is extracting pictorially independent information about the face.

Method

Participants

Fifteen participants (7 females and 8 males) aged between 19 and 22 years (M = 20.667 years, SD = 0.816) took part in the experiment. All had normal or corrected to normal vision. None had previous experience of the faces or had taken part in Experiment 1.

Materials

The materials were the same faces used in Experiment 1. An additional four identities were added to the set, which were prepared in the same way as the original 20 individuals from Experiment 1. The 24 identities were split into two sets of 12 people

each. Three images were used per individual: morphs representing 95%/5%, 50%/50% and 5%/95% mixes (i.e. pseudo-ages of 20, 40 and 60) of the 20 year old and 60 year old images respectively.

Apparatus

The apparatus used in Experiment 1 was employed.

Procedure

The procedure was similar to that used in Experiment 1. All participants received three phases: a first presentation phase, a training phase and a test phase.

Presentation phase

Images of twelve people were shown, four from the 20-year-old image and four from the 60-year-old image. The final four were to be learnt from two images. However, in this phase, only one of those images was shown. Of these four people, two were randomly selected to be shown from the 20-year-old image whilst the other two were shown from the 60-year-old image. Each face was presented for 5 seconds each with 0.5 seconds between each face and participants were instructed to try to learn the name/face pairings.

Training phase

In the training phase, all 16 images were used. In the first part of training, the 16 images were split into four blocks containing four faces each. Within each block, all images showed faces of the same age (20 or 60). As in Experiment 1, to progress from one block to the next, and to complete the first part of training, participants had to correctly name all four people in the block, without making an error, on three separate occasions. The

second part of training presented participants with all 16 images to be named. During this part, only the 12 possible names were given as options (i.e. there were no duplicate names for the faces learnt from two images). As in the previous experiment, participants had to correctly name all 16 images, without making an error, on three separate occasions.

Test phase

The test phase consisted of three blocks. Within each block, participants were presented with 24 faces (12 targets/12 distractors). All of the faces within a block were of the same pseudo-age (either 20, 40 or 60 years). The presentation order of these three blocks was counterbalanced across participants. Responses were made by clicking on one of two buttons located underneath the face on the screen. Participants were instructed to respond *Yes* if they thought they recognised the individual or *No* if they did not.

Results

The mean proportion correct scores for each of the three learning conditions across the three test conditions are shown in Figure 3.

[Insert Figure 3 here]

The mean proportion correct scores were entered into a two-way repeated measures ANOVA with learnt age and test age as independent factors. There were significant main effects of learnt age, with a large effect size: $F(1.9, 26.0) = 64.290$, $MSE = 0.035$, $\eta_G^2 = .418$, $p < .001$, and of test age, with a smaller effect: $F(1.5, 20.5) = 4.485$, $MSE = 0.044$,

η_G^2 = .05, p = .020. A significant interaction between learnt age and test age with a large effect size was also found: $F(2.8,39.8)$ = 72.230, MSE = 0.053, η_G^2 = .65, $p < .001$. Post-hoc analyses revealed that learning only the 20-year-old image resulted in a poorer recognition the older the face became (20 vs. 40: $p < .001$, 40 vs. 60: $p < .001$). A similar, but inverse, linear trend was also observed for the learning of the 60-year-old image (60 vs. 40: $p < .001$, 40 vs. 20: $p < .001$). When both the 20 and 60-year-old images were learnt there was significantly poorer recognition performance of the unseen 40-year-old image in comparison to either of the learnt ages (both p's < .001).

To analyse the significant interaction, a simple main effects analysis was conducted. Following from Experiment 1, testing across the three ages produced expected generalisation gradients for both the learn 20-year-old ($F(2.0, 28.0)$ = 77.132, MSE = 0.026, η_G^2 = .707, $p < .001$) and learn 60-year-old images ($F(2.0, 28.0)$ = 41.245, MSE = 0.043, η_G^2 = .626, $p < .001$). In addition, learning both the 20-year-old and 60-year-old images also resulted in a significant difference for the three test ages ($F(1.0, 14.0)$ = 21.000, MSE = 0.007, η_G^2 = .500, $p < .001$). Post-hoc analyses revealed that testing with the 40-year-old image resulted in lower accuracy than testing with either the 20-year-old or 60-year-old images (both p's < .001). Accuracy on the 20-year-old and 60-year-old images was identical.

Recognition accuracy on the 20-year-old test image across the three learning conditions yielded a significant difference ($F(1.0, 14.0)$ = 97.715, MSE = 0.024, η_G^2 = .823, $p < .001$). This effect was driven by good recognition after learning the 20-year-old image (either individually or in conjunction with the 60-year-old image) in comparison to poor

performance after learning the 60-year-old image only (both p's < .001). Performance after learning either the 20-year-old view only or both the 20- and 60-year-old images was identical. Recognition accuracy on the 60-year-old test image across the three learning conditions also yielded a significant difference (F(1.0, 14.0) = 142.316, MSE = 0.019, η_G^2 = .871, p < .001). Similar to the previous result, this effect was driven by good recognition after learning the 60-year-old image (either individually or in conjunction with the 20-year-old image) in comparison to poor performance after learning the 20-year-old image only (both p's < .001). Performance after learning either the 60-year-old view only or both the 20- and 60-year-old images was identical.

The critical comparison is the performance obtained for the 40-year-old image, as it was novel and previously unseen. Comparing performance on the 40-year-old image across the three learning conditions yielded a significant difference (F(2.0,28.0) = 8.034, MSE = 0.033, η_G^2 = .188, p = .002). Comparisons across the learning conditions revealed that learning both ages produced better recognition accuracy than learning either of the single ages (20 vs. both: p = .003, 60 vs. both = .041). There was no difference in either of the single age learning conditions (p > .05).

As in Experiment 1, combinations of learn and test views in which pseudo-age differed markedly proved challenging. Participant performance was compared against chance level (proportion correct = 0.5) using a one-sample t-test for each learn/test view combination and the results are shown in Table 3.

[Insert Table 3 here]

Discussion

Experiment 2 revealed that learning two different images (a pseudo 20-year-old image and a pseudo 60-year-old image) of an individual's face does lead to better recognition accuracy on a novel image that is a 50/50 morph of the learnt images compared to learning only one image. Therefore, the results of this experiment suggest that learning a face at two different ages allows for the abstraction of structural cues to identity. Importantly however, performance after learning two views of the face still falls statistically short of that obtained for the recognition of either of the originally studied exemplars and thus an alternative explanation is possible; the increase in accuracy observed after studying two images is the result of an additive effect. Such an account would imply a pictorially dependent recognition system as opposed to a structural mechanism for face recognition as the latter would be expected to result in equal performance on the novel image. This possibility was addressed in Experiment 3.

Experiment 3

The results of Experiment 2 suggest that learning two views of a face does lead to better recognition of a novel unseen view that falls between the learned views. This result could suggest the formation of a crude abstract structural representation or is possibly the result of a recognition process based upon interpolation between stored views (Bulthoff & Edelman, 1992; Tarr & Bulthoff, 1998; Wallraven, Schwaninger, Schuhmacher & Bultoff, 2002). To examine this further, Experiment 3 investigated whether learning two views of a face leads to better extrapolation to a novel view of a face that lies beyond the learnt images. If abstract structural codes are being formed after learning two views then extrapolation to a new image that falls beyond the

boundaries of the studied images should be greater than if a single view is learned as common elements of the face that determine identity that make up the structural code can be extracted from multiple sources.

Method

Participants

Seventeen participants (15 females and 2 males) aged between 19 and 46 years (M = 21.882 years, SD = 6.963) took part in the experiment. All had normal or corrected to normal vision. None had previous experience of the faces or had taken part in either of the previous experiments.

Materials and Apparatus

The same materials and apparatus from Experiment 2 were used.

Procedure

The procedure was the same as that of Experiment 2 except for the 40-year-old image replaced the 60-year-old image during the training phase.

Results

The mean proportion correct scores for each of the three learning conditions across the three test conditions are shown in Figure 4.

[Insert Figure 4 here]

The mean proportion correct scores were entered into a two-way repeated measures ANOVA with learned age and test age as independent factors. There were significant main effects of learned age, with a large effect size: $F(1.5,23.9)$ = 42.574, MSE = 0.041,

η_G^2 = .331, p < .001, and of test age, also with a large effect: $F(2.0,32.0)$ = 81.694, MSE =

0.039, η_G^2 = .554, p = .027. A significant interaction between learned age and test age

with a large effect size was also found: $F(3.5, 56.0)$ = 47.610, MSE = 0.036, η_G^2 = .529, p

< .001.

To analyse the interaction, a simple main effects analysis was conducted. In line with

Experiments 1 and 2, generalisation gradients were found for the three learning

conditions. Recognition accuracy for faces learnt from the 20-year-old image differed

across the three test images ($F(1.3,20.8)$ = 139.649, MSE = 0.015, η_G^2 = .846, p < .001).

Likewise learning the 40-year-old image only produced a difference across the three

test images ($F(1.9,31.2)$ = 19.120, MSE = 0.033, η_G^2 = .420, p < .001). In this condition,

the 40-year-old test image was recognised significantly better than either the 20-year-

old or 60-year-old images (p < .001) whilst the latter two test conditions did not differ

(p > .05). Finally, learning both the 20-year-old and the 40-year-old image also revealed

a difference across test images ($F(1.0,16.0)$ = 77.714, MSE = 0.010, η_G^2 = .764, p < .001).

In this learning condition accuracy on the 60-year-old image was significantly lower

than that of the 20-year-old or 40-year-old images (both p's < .001). The latter two test

conditions yielded identical accuracy.

Performance on the 20-year-old test image across the three learning conditions yielded

a significant difference ($F(1.0, 16.0)$ = 48.944, MSE = 0.016, η_G^2 = .671, p < .001). This

effect was driven by poor recognition after learning the 40-year-old image in

comparison to good performance after learning the 20-year-old (p < .001) or both the

20- and 40-year old images (p < .001). Performance after learning either the 20-year-old

view only or both the 20- and 40-year-old images was identical. Testing with the 40-year-old image also revealed a significant difference across learning conditions (F(1.0, 31.0) = 60.870, MSE = 0.014, η_G^2 = .717, p < .001. In this test condition the learn 20-year-old image condition yielding significantly lower accuracy than either learn 40-year old (p < .001) and both 20- and 40-year-old conditions (p < .001). The 40-year-old and two view learning conditions resulted in identical performance.

Testing with the 60-year-old image also led to a significant difference between learning conditions (F(1.6, 26.1) = 35.218, MSE = 0.024, η_G^2 = .510, p < .001). Pairwise comparisons revealed a significant difference between learning the 20-year-old image only with both learning the 40-year-old image only (p < .001) and two view learning (p < .001). Recognition accuracy after learning the 40-year-old image only and both the 20- and 40-year-old images did not differ (p = .572). Participant performance was also compared against chance level of 0.5 for each learn/test view combination. The results are shown in Table 4.

[Insert Table 4 here]

Discussion

The results of Experiment 3 reveal that learning two views of a face does not lead to an enhanced ability to extrapolate to a novel view of the face in comparison to learning a single view. The data appear to suggest that recognition of the 60-year-old image was dependent upon generalization from the 40-year-old image, as opposed to any generalized abstraction of the 20-year-old and 40-year-old images. Taken together, Experiments 2 and 3 suggest image-dependency in the recognition of newly learned

faces such that recognition success appears to be related to the ability to generalize from a previously seen image of a face. The level of recognition with respect to chance performance also appears similar to Experiment 2 in which generalisation to, or from, the 40-year-old image produces accuracy similar to that of chance. It should be noted however that as conditions involving extreme changes in both Experiments 2 and 3 yield significantly lower than chance performance it is unlikely that participants are simply guessing when generalising to or from the 40-year-old image.

General Discussion

 The experiments presented provide further support for an image dependent system for the recognition of faces learnt from a limited number of images. Experiment 1 examined the extent to which faces learnt from a single image can be generalised across different pseudo-ages. It was found that generalisation was possible for a limited range equating to approximately 20% morph away from the learnt image. The performance obtained after a change of more than 20% was lower than 25% accuracy for all learning conditions, suggesting that there are strict limits to the ability of the generalisation processes to generalise to novel views of a face. Despite this large drop after morph changes of greater than 20%, recognition appears to follow a generalisation gradient the more the face is changed from the learnt view, suggesting a pictorially dependent recognition system. Figure 2 does reveal however that within the 20% morph range generalisation is good such that performance is not significantly different from the recognition rates for the learnt image. A very similar result was found by Longmore et al. (2008, Experiment 1) in which newly learnt faces were successfully recognised after a change in head orientation of 17º. These results suggest that within tolerance limits, the face generalisation system is highly effective.

Since only one image of each face was learnt in Experiment 1, it might not be considered surprising that recognition was likely to be mediated by a single picture. Experiment 2 extended this finding to see whether learning two images of a face that differed in apparent age aided recognition of a previously unseen image. It was found that learning two images of a face did lead to better recognition of the novel third image (depicting the face at an intermediate age) but this level of performance still fell significantly below that obtained for the studied images. At first glance, it would appear that learning a face from two different ages does lead to the formation of abstract structural codes. There are two possible explanations why this might not be the case however. First, one possible structural mechanism that would make a similar prediction is the facial prototype effect. As the novel testing image was a 50/50 morph of the learnt images, it can be considered to act as the prototype for the two learnt images. So, on the basis of previous work (Bruce, Doyle, Dench, & Burton, 1991; Cabeza, Bruce, Kato, & Oda, 1999), we might expect that recognition performance on this prototype image would be as good as, or very similar to, the level for the learnt images. However, our results indicated that recognition on this novel image was poorer than that obtained for the learnt images, suggesting that recognition was likely to be based upon generalisation from the seen exemplars.

A second possible reason for this lies in this categorical perception of the faces. It has been shown facial expressions (Calder, Young, Perrett, Etcoff, & Rowland, 1996), sex (Campanella, Chrysochoos, & Bruyer, 2001) and race (Levin & Angelone, 2002) are judged on a categorical basis. That is, rather than being viewed as a continuum between different expressions or races, faces appear to suddenly change from one expression or

race to another. Such demonstrations of this effect utilise a morphing procedure similar to that use in the experiments reported here. As there were five images used to depict different ages, it is likely that the middle image (i.e. the 40-year-old image) was likely to be perceived as being more like either the 20 or the 60-year-old image. Due to the range of images it is possible that for some identities, the middle image was more like the 20-year-old picture whilst for others, the 60-year-old was more similar. It might not therefore be considered surprising that learning both a 20-year-old and a 60-year-old image was advantageous as either image could be used for generalisation. To test this possibility, Experiment 3 examined whether extrapolation from learnt views to a novel image beyond the range of the learnt views was possible. If recognition was mediated by structural coding then it would be expected that extrapolation to the novel view after learning two views should be more successful than if a single image had been learnt. However, the results clearly indicated that learning two views did not provide any advantage over learning a single view and it appears that the recognition of the novel 60-year-old image was being driven by generalisation from the 40-year-old image (i.e. the closest image pictorially). As noted in the discussion of Experiment 1 however, it does appear that varying levels of generalisation to a novel view of a face is possible and it is only in extreme changes (i.e. the 20-year-old and 60-year-old images) that recognition accuracy falls close to 0.

The differences in generalisation performance found in Experiments 2 and 3 might also be related to the finding that interpolation between images of faces is easier than extrapolation (Wallraven et al., 2002), possibly due to the higher complexity of extrapolation in comparison to interpolation (Ullman & Basri, 1991). Wallraven et al. (2002) presented participants with two views of a face that differed in head orientation

and tested recognition for the same view, a head orientation that fell between the studied views within the horizontal plane (interpolation), beyond the studied views in the horizontal plane (extrapolation) or a change in the vertical plane. The results demonstrated that participants were more successful at identifying faces that required interpolation between the studied views than they were at identifying faces that required extrapolation. These results were found with a change in head orientation and Experiments 2 and 3 extend this result to demonstrate that the relative difficulties faced by the face recognition processes for interpolation and extrapolation appear to apply to image variation due to age.

It is important to note though that in both Experiments 2 and 3, participants learnt two separate images of an individual and the only indication that the two images did indeed depict the same individual was the fact that they were given the same name. Menon, White and Kemp (2015) have reported that informing participants that two different images of an individual depict the same person can improve generalisation to a novel, third image of that person whilst telling participants that the two images depict different people can reduce recognition accuracy of the third image. The results obtained in Experiments 2 and 3 may therefore be indicative of participants treating the two learnt images as two different individuals.

Overall, the results reported suggest that the recognition of faces learnt from a limited number of sources operates on pictorial codes and that recognition is dependent upon how similar a test image is to the stored representation of that face. Increasing the number of images presented is likely to increase the chances of successful identification of an unfamiliar face (c.f. Bindemann & Sandford, 2011). However it is for future

research to determine how many images are needed to elicit familiar face processing like performance.

Acknowledgements

References

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods, 37*(3), 379-384. doi:10.3758/BF03192707

Bindemann, M., Avetisyan, M., & Blackwell, K. A. (2010). Finding needles in haystacks: Identity mismatch frequency and facial identity verification. *Journal of Experimental Psychology: Applied, 16*(4), 378-386. doi:10.1037/a0021893

Bindemann, M., & Sandford, A. (2011). Me, myself, and I: Different recognition rates for three photo-IDs of the same person. *Perception, 40*(5), 625-627. doi:10.1068/p7008

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*, 105-116.

Bruce, V., Doyle, T., Dench, N., & Burton, M. (1991). Remembering facial configurations. *Cognition, 38*(2), 109-144. doi:10.1016/0010-0277(91)90049-A

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

Bruck, M., Cavanagh, P., & Ceci, S. J. (1991). Fortysomething: Recognizing faces at ones 25th reunion. *Memory & Cognition, 19*(3), 221-228. doi:10.3758/BF03211146

Bulthoff, H. H., & Edelman, S. (1992). Psychophysical support for a 2-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences of the United States of America, 89*(1), 60-64. doi:10.1073/pnas.89.1.60

Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow Face Matching Test. *Behavior Research Methods, 42*(1), 286-291. doi:10.3758/BRM.42.1.286

Cabeza, R., Bruce, V., Kato, T., & Oda, M. (1999). The prototype effect in face recognition: Extension and limits. *Memory & Cognition, 27*(1), 139-151. doi:10.3758/BF03201220

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

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioural evidence and the face-space metaphor. *Visual Cognition, 8*(2), 237-262.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42-45.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62*(2), 201-222. doi:10.1016/S0010-0277(96)00785-8

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31*(10), 1221-1240. doi:10.1068/p3252

Jenkins, R., White, D., Van Montfort, X., & Burton, A. M. (2011). Variability in photos of the same face. *Cognition, 121*(3), 313-323. doi:10.1016/J.Cognition.2011.08.001

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17*(5), 577-596. doi:10.1080/09658210902976969

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology, 11*(3), 211-222.

Krouse, F. L. (1981). Effects of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology, 66*(5), 651-654. doi:10.1037/0021-9010.66.5.651

Levin, D. T., & Angelone, B. L. (2002). Categorical perception of race. *Perception, 31*(5), 567-578. doi:10.1068/p3315

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception and Performance, 34*(1), 77-100. doi:10.1037/0096-1523.34.1.77

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shape. *Philosopghical Transactions of the Royal Society of London: Series B, 200*, 269-294. doi:10.1098/rspb.1978.0020

Menon, N., White, D., & Kemp, R. I. (2015). Identity-level representations affect unfamiliar face matching performance in sequential but not simultaneous tasks. *Quarterly Journal of Experimental Psychology, 68*(9), 1777-1793. doi: 10.1080/17470218.2014.990468

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4*(2), 61-64.

Norrkross. (2009). MorphX (Version 2.9.5). Retrieved from http://www.norrkross.com/software/morphx/morphx.php

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3*(4), 406-417. doi:10.1037//0278-7393.3.4.406

Tarr, M. J., & Bulthoff, H. H. (1998). Image-based object recognition in man, monkey and machine. *Cognition, 67*(1-2), 1-20. doi:10.1016/S0010-0277(98)00026-2

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*, 992-1006.

Wallraven, C., Schwaninger, A., Schuhmacher, S., & Bulthoff, H. H. (2002). View-based recognition of faces in man and machine: Re-visiting inter-extra-ortho. *Biologically Motivated Computer Vision, Proceedings, 2525*, 651-660. WOS:000181471800065

Wells, G. L. (1993). What do we know about eyewitness identification. *American Psychologist, 48*(5), 553-571. doi:10.1037//0003-066x.48.5.553

White, D., Dunn, J. D., Schmid, A. C., & Kemp, R. I. (2015). Error Rates in Users of Automatic Face Recognition Software. *Plos One, 10*(10), e0139827. doi:10.1371/journal.pone.0139827

White, D., Kemp, R. I., Jenkins, R., Matheson, M., & Burton, A. M. (2014). Passport officers' errors in face matching. *Plos One, 9*(8), e103510. doi:10.1371/journal.pone.0103510

Table 1: Mean proportion correct obtained within each of the five test blocks

	Block 1	Block 2	Block 3	Block 4	Block 5
Mean proportion correct	0.806	0.784	0.788	0.791	0.741

Table 2: Performance obtained within each learn/test view combination of Experiment 1 versus chance level performance of a mean proportion correct score of 0.5 using a one-sample t-test. NS = not significantly difference from chance, S↑ = performance was significantly above chance, S↓ = performance was significantly below chance, ** = Performance for all participants was 100% (and thus above chance) however with no variance within the condition the analysis is not possible.

		Learn Age				
		20	30	40	50	60
Test Age	20	**	S↑	S↓	S↓	S↓
	30	S↑	**	S↑	S↓	S↓
	40	NS	S↑	**	S↑	NS
	50	S↓	S↓	S↑	S↑	S↑
	60	S↓	S↓	NS	S↑	**

Table 3: Performance obtained within each learn/test view combination of Experiment 2 versus chance level performance of a mean proportion correct score of 0.5. NS = not significantly difference from chance, S↑ = performance was significantly above chance, S↓ = performance was significantly below chance, ** = Performance for all participants was 100% (and thus above chance) however with no variance within the condition the analysis is not possible.

		Learn Age		
		20	60	Both 20+60
	20	**	S↓	**
Test Age	40	NS	NS	S↑
	60	S↓	**	**

Table 4: Performance obtained within each learn/test view combination of Experiment 3 versus chance level performance of a mean proportion correct score of 0.5. NS = not significantly difference from chance, S↓ = performance was significantly below chance, ** = Performance for all participants was 100% (and thus above chance) however with no variance within the condition the analysis is not possible.

		Learn Age		
		20	40	Both 20+40
	20	**	NS	**
Test Age	40	NS	**	**
	60	S↓	NS	NS

Figure 1. Example of the images used in Experiment 1. The pseudo-ages of the faces were (from left to right): 20, 30, 40, 50, and 60 years.
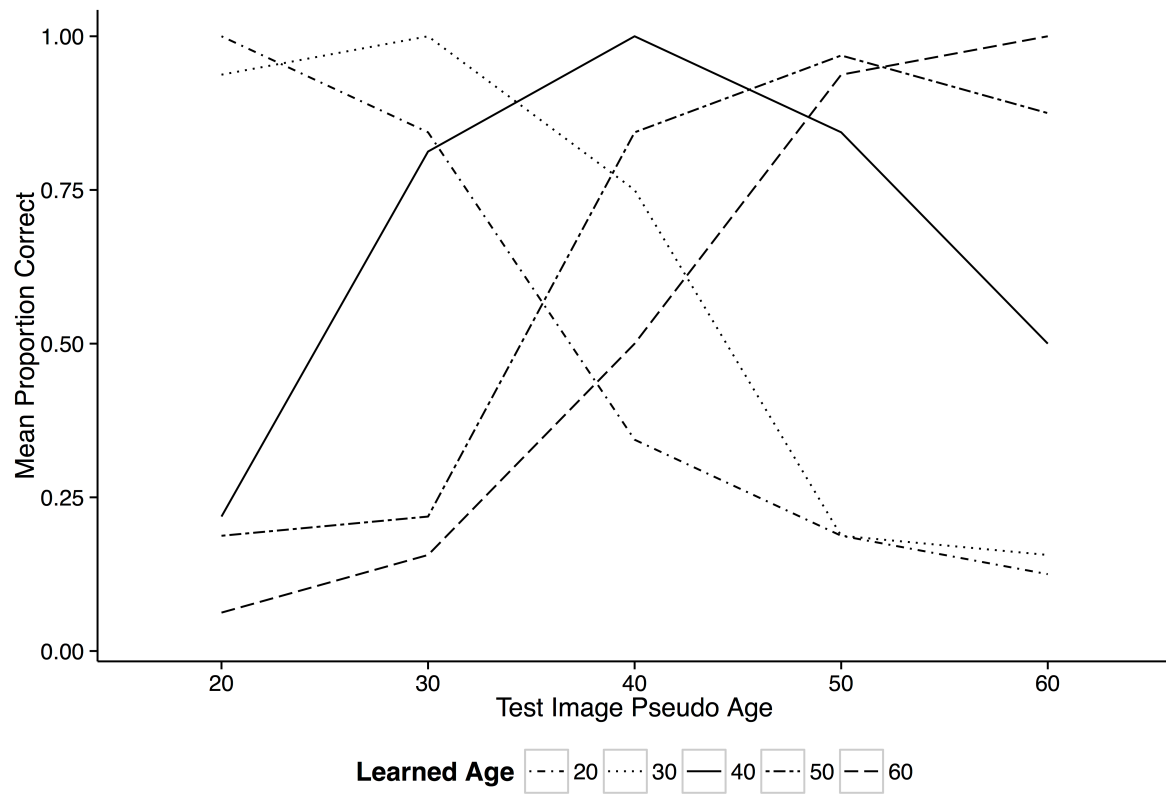
Figure 2. Mean proportion correct scores for the five learnt pseudo-ages across the five possible testing ages in Experiment 1.
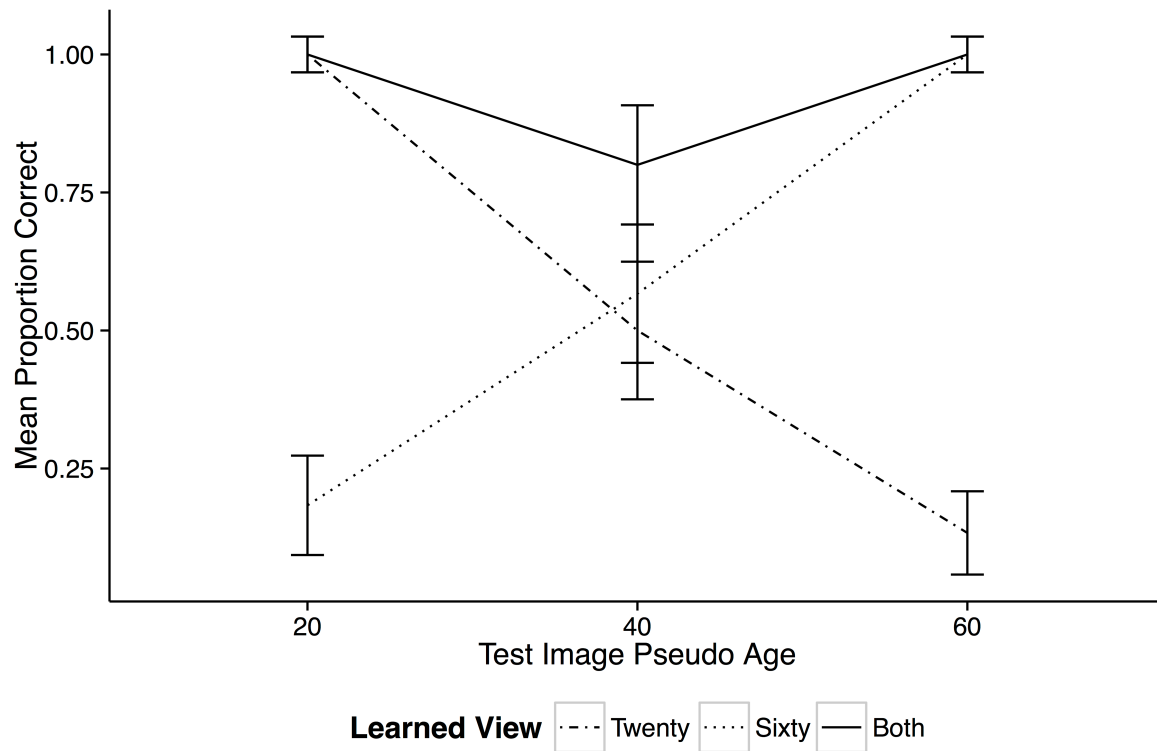
Figure 3. Mean proportion correct recognition scores after single and dual pseudo-age learning on pseudo-ages that were seen or unseen during learning in Experiment 2. Error bars represent 95% confidence intervals (Cousineau, 2005) with Morey (2008) corrections.
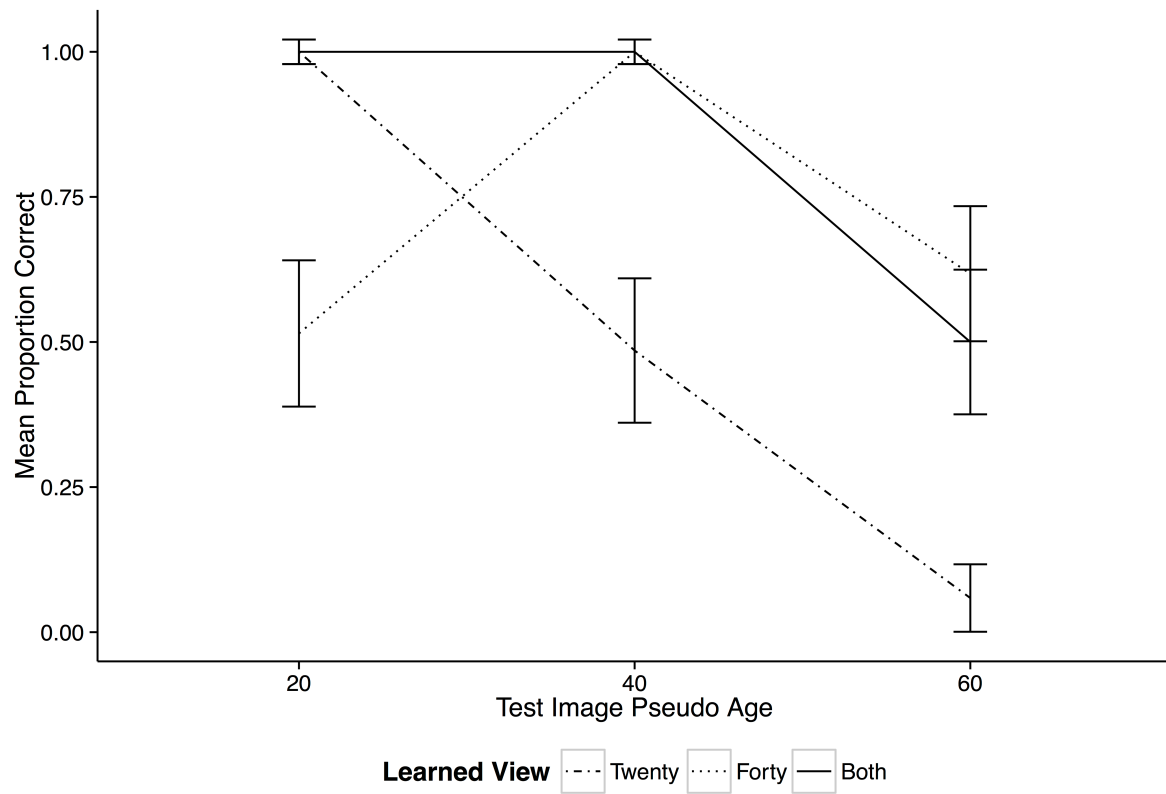
Figure 4. Mean proportion correct recognition scores after single and dual pseudo-age learning on pseudo-ages that were seen or unseen during learning in Experiment 3. Error bars represent 95% confidence intervals (Cousineau, 2005) with Morey (Morey, 2008) corrections.