The utility of multiple synthesized views in the recognition of unfamiliar faces

Scott P. Jones, Dominic M. Dwyer & Michael B. Lewis

To cite this article: Scott P. Jones, Dominic M. Dwyer & Michael B. Lewis (2017) The utility of multiple synthesized views in the recognition of unfamiliar faces, The Quarterly Journal of Experimental Psychology, 70:5, 906-918, DOI: 10.1080/17470218.2016.1158302

To link to this article: http://dx.doi.org/10.1080/17470218.2016.1158302

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Accepted author version posted online: 24 Feb 2016.
Published online: 24 Mar 2016.

Submit your article to this journal

Article views: 416

View related articles

View Crossmark data

Citing articles: 1 View citing articles
The ability to accurately identify an unfamiliar individual based on a single photograph is often poor and prone to error. A number of studies have reported that even when provided with the most helpful conditions (e.g., photographs taken on the same day in relatively similar conditions), people are remarkably poor at identifying the target individual (Bruce, Henderson, Newman, & Burton, 2001; Clutterbuck & Johnston, 2002; Megreya & Bindemann, 2009; Megreya & Burton, 2006). This fallibility extends outside the laboratory as demonstrated by Kemp, Towell, and Pike (1997) who found that supermarket staff failed more than 50% of the time to notice that customers had presented photo identification that did not depict that individual—despite being informed that they were under observation and taking part in a study (see also, Davis & Valentine, 2009; Megreya & Burton, 2008; White, Kemp, Jenkins, Matheson, & Burton, 2014). In contrast, our capacity for identifying familiar individuals is markedly superior. People excel at recognizing familiar faces even when they are shown poor-quality images, are given limited inspection time, or are asked to recognize individuals whom they have not seen for a long period of time (Bruck, Cavanagh, & Ceci, 1991; Burton, Wilson, Cowan, & Bruce, 1999; Buttle & Raymond, 2003). For example, Bahrick, Bahrick, and Wittlinger (1975) observed a 90% correct matching rate for names and faces of high-school graduates asked to recognize members from their yearbook photo over 15 years since their last encounter.

As a result of the disparity in performance, much attention has been given to the effects of familiarity in face perception (for reviews see, Hancock, Bruce,
For an updated version of the model using a similar method of construction see the Basel face model (Paysan, Knothe, Amberg, Romdhani, & Vetter, 2009). The update includes better original scans using a more advanced scanning device and an improved correspondence algorithm. However, the method of construction remains almost identical to the that of previous versions.
which represents deviation from this given 3D and texture average (for a more in-depth discussion on how the two kinds of information are computed, see Blanz & Vetter, 1999; Vetter & Poggio, 1997). Principal component analysis (PCA) is then conducted to find the eigenvectors allowing a new range of faces to be synthesized. What this method of construction (and others like it) enables is for each face to be rendered under clearly defined lighting conditions or views (for a detailed discussion of PCA and generated faces, see Hancock, Burton, & Bruce, 1996; Vetter & Walker, 2012).

Most critically for our current concerns, these methods for face modelling also allow a single 2D photographic input to be semi-automatically reconstructed—in essence, creating a 3D representation of that individual. This derived representation of the individual can then be manipulated within the computer program to extract multiple views of the face. Early work has found that people will accurately match between original and reconstructed images (albeit not as accurately as between different original images), suggesting that reconstructed images are a reasonably faithful reflection of the original (Bailenson, Beall, Blascovich, & Rex, 2004).

More recently it has been demonstrated that performance, in an old/new recognition task, can be enhanced by the presence of computer modelled faces created from a single 2D photograph input compared to an original single image (Liu, Chai, Shan, Honma, & Osada, 2009). Experiment 1 compared recognition performance of a single image (control) to that of the same image, plus extra synthesized views (experimental condition) with better performance in the experimental condition. Subsequent experiments demonstrated better performance following presentation of either still or dynamic multiple generated faces or user-controlled exposure to a 3D generated bust. While these studies certainly suggest that generated images can improve recognition compared to exposure to a single original image, there are some potential caveats that limit the conclusions that can be drawn. In all experiments, the number of images presented in the control condition did not match the total number of images presented in the experimental condition, and there was explicit instruction regarding the use of synthetic images, which may have directly influenced performance in the experimental conditions. Moreover, the single control images themselves consisted of “135 laser-scanned models and their texture maps” that resembled, but were not actual, photographs (see, Liu et al., 2009, p. 993). Thus, while these studies clearly show some ability for people to learn about artificially generated faces, they do not conclusively demonstrate that using such generated faces can be used to support superior performance to that produced by limited real images.

In summary, people clearly generalize between real photographic and computer synthesized face images, and there is also evidence that they can learn to recognize individuals through exposure to generated images. What remains to be assessed is whether such generated images—derived from a single original photograph—can be used to support identification performance that is superior to that produced by exposure to the original photograph alone. Therefore, in the current two experiments, we used commercially available software (SI FaceGen Modeller), to generate multiple images—derived from single original source photographs—to further explore the utility of artificial face synthesis as a means of facilitating recognition within a sequential face matching procedure.

**Experiment 1**

While providing more face information through multiple images helps improve recognition (Dowsett, Sandford, & Burton, 2016; White, Burton, Jenkins, & Kemp, 2014), often only limited real images are available (e.g., a single wanted picture, CCTV image, or passport photo). The SI FaceGen Modeller, used in the following studies, requires (at least) a single front-view input to model a face. The software is capable of modelling an input face onto a photogrammetric bust. From this, the bust can be manipulated and captured at different yaw and pitch rotations. Experiment 1 compared the effects of these synthesized views with original photographic images displaying different rotations of a target individual. Because the processing of unfamiliar faces is negatively affected by changes in viewpoint, we examined whether exposure to these multiple views aided identification performance in a sequential matching task that involved a viewpoint change: Participants were presented with a line-up procedure that required selecting a target from a test array of images presented at a novel angle (Table 1 summarizes the design, and Figure 1 shows an example of the test array). Participants were given exposure to target individuals under four conditions: original image—only
the original front-on image; test image—the original front-on image plus a generated view at the novel test angle; photographic views—the original front-on image plus additional original images at different angles (these did not include the novel angle used in the test arrays to prevent direct matching of images between exposure and test); synthesized views—the original front-on image plus additional synthesized images at different angles (including the novel test angle). Comparing the photographic views with the original image conditions establishes the degree of improvement produced by the addition of multiple real images. The test image and synthesized views conditions establish the degree of improvement produced by adding either a single generated image at the test angle or multiple generated images across a range of angles including the test angle.

**Method**

**Participants**

Twenty-four students, aged 18–24 years, were recruited from Cardiff University and completed the experiment in return for payment of £2. All participants reported normal or corrected-to-normal vision.

**Stimuli, face synthesis, and test arrays**

The stimuli comprised eight face images selected from the CMU Multi-Pie database (Gross, Matthews, Cohn, Kanade, & Baker, 2010) to become target individuals. Each image was taken under the same lighting conditions with the individual facing the camera. Each individual was between the ages of 19 and 45 years and was chosen to avoid the presence of non-face cues (e.g., glasses and excessive facial hair) that could obscure any featural or structural information about the face, or hinder the generation of synthesized faces. Along with the front-view photo, six additional photos of each target were chosen: a total of 48 faces. The six additional photos panned from the left-side view to a right-side view with 30° increments between each photo (see Figure 2, Panel A).

Synthesized faces were generated using SI FaceGen Modeller 3.1 (developed by Singular Inversions, Toronto, Canada). The SI FaceGen Modeller 3.1 is a face-generating, 3D modelling, software similar to the three-dimensional morphable face model (Blanz & Vetter, 1999, 2003). FaceGen was created on the basis of 273 laser scans of a range of individual faces. Using a single front-view photo, the software is able to synthesize a computational representation of the face.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Initial exposure</th>
<th>Training sequence</th>
<th>Test arrays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original image</td>
<td>0°</td>
<td>+, +, +, 0°, +, +</td>
<td>Five 30°L faces</td>
</tr>
<tr>
<td>Test image</td>
<td>0°</td>
<td>+, +, *30°L, 0°, +, +</td>
<td>Five 30°L faces</td>
</tr>
<tr>
<td>Photographic views</td>
<td>0°</td>
<td>90°L, 60°L, +, 0°, 30°R, 60°R, 90°R</td>
<td>Five 30°L faces</td>
</tr>
</tbody>
</table>

Note: Training sequence presented in order. All conditions began with the target front-view photo. The training sequence was different for each condition but included fixation crosses “+” and images of the target face at various angles of yaw (indicated by a numerical angle and an indication of whether this was to the right, R, or left, L). The “*” indicates that the image was synthesized using FaceGen rather than being photographic.

![Figure 1](image-url)
of a face on a rotatable 3D model. The best “fit” for shape and colour parameters (along with position and lighting parameters) is selected until the model matches the face region of the “photo”. The most likely 3D face that produced the photo according to the model is the resulting output. For the user, this is accomplished by placing landmarks upon the key features of the face (e.g., corners of the mouth, jaw line, eyes). In the current experiment, the software was used to synthesize still 2D images of a target face, as if they were captured at a variety of angles. This resulted in the creation of counterpart synthesized versions of the targets, and 2D stills were taken akin to the angles of the photo stimuli selected from the face database (see Figure 2, Panel B). Each photo and synthesized image was cropped to remove hair, using Adobe Photoshop 6, and was displayed on a black background on screen at 600 × 463 pixels, subtending to an approximate visual angle of 23.4° × 18.4°.

Test arrays were made by cropping photographs of the target individual and four other individuals (foils) that were displayed at a 30° angle (facing left when looking at the screen) within the array (see Figure 1). The five faces were presented simultaneously with the target face in a random position within the array. There were no target-absent arrays. The foils were chosen on the basis that they matched a basic verbal description of the target face. The target in Figure 2 for example is, “a male with short dark hair”. Each array was homogenized so that the sizes of each array were identical (800 × 267 pixels, subtending an approximate visual angle of 29.9° × 10.9°). All arrays were subject to a tonal change using the brightness/contrast adjustment tool in Adobe Photoshop image editing software. Each array was adjusted to +50% brightness and −20% contrast of the original image.

**Design and procedure**

Each participant completed four conditions (original image, test image, synthesized views, and photographic views) as part of a within-subjects sequential matching task. Before each condition began, a front-view photo depicting a target individual was displayed. In the synthesized views condition, exposure comprised multiple computer-generated views that were presented either side of the original target. These began at 90° left and presented the following angles: 60° left, 30° left, 30° right, 60° right, and 90° right. This progression was intersected by the original image that was presented between the two 30° images. The photographic views gave the same exposures as the synthesized condition, but used original photos at all orientations. The exception in this condition was that the 30° left-facing image was removed and was replaced with a fixation cross in order to prevent direct matching to the images used in the test array. Presentations in the multiple exposure conditions (synthesized and photographic views) always ran from 90° left to 90° right and

![Figure 2. Examples of stimuli used in Experiments 1 and 2. Panel A represents the photographic views used in Experiment 1. Panel B represents the counterpart synthesized images used in the synthesized views condition (Experiments 1 and 2). All synthesized examples were created from a single front on input.](image-url)
returned, such that the last presentation before test was a left-facing 90° profile view. The original image condition involved only the presentation of the front on photograph images, with the generated faces replaced by fixation crosses to maintain the overall timing across the trial. The test image condition displayed the front-view photos consistent with the previous condition, and an additional synthesized image displayed at a left 30° angle. This condition also utilized fixation crosses like the original image condition (Table 1 summarizes the design). Including the face presented before each condition, all face stimuli and fixation crosses were presented for 2 s with an inter-stimulus interval (ISI) of 1 s between each stimulus presentation. The total presentation time was 45 s (i.e., a total of 30 s for all stimuli presentations plus a total of 15 s for the ISI). Every participant was tested on two different face stimuli within each experimental condition. All conditions were run in blocks, but counterbalanced such that each condition was presented first, second, third, or fourth equally often, and this was rotated such that every condition was placed in every presentation order across the counterbalance. Similarly, every face appeared within each condition across the counterbalance. At test, participants were asked to select the target seen during exposure from an array of five faces, in which the target was always present. Response time was limited to 10 s. Participants keyed response via the keyboard, using the numbers 1–5, which corresponded to the numbers below each face. Following this, they were asked to make a judgement about how confident they were about their choice using a button response to a 7-point Likert scale (1: “Not at all confident”, 7: “Extremely confident”).

**Results and discussion**

Figure 3A displays identification accuracy as percentage correct for each condition (test image, original image, synthesized views, and photographic views). Exposure to the synthesized and photographic views conditions resulted in better performance than that to the brief controls (test image, original image). There was little apparent difference between the synthesized and photographic views conditions, and little difference between the two brief exposure control conditions. A within-subjects analysis of variance (ANOVA) failed to observe a main effect of condition, $F(3, 69) = 2.28$, $MSE = 0.148$, $p = .087$. However, planned comparisons revealed that higher accuracy was observed for the synthesized views condition than for the original image exposure, $F(1, 23) = 9.47$, $MSE = 0.510$, $p = .005$, and the test image exposure conditions, $F(1, 23) = 4.28$, $MSE = 0.510$, $p = .049$. No other differences were observed between the conditions (largest, $F(1, 23) = 2.76$, $MSE = 0.375$, $p = .110$, between photographic views and test image condition).

Figure 3B displays the CA data. Again, performance in the synthesized and photographic views conditions was better than that in the test and original image conditions. Analysis of the CA scores revealed a main effect of condition, $F(3, 69) = 3.30$, $MSE = 13.044$, $p = .025$. Planned comparisons revealed that training in the photographic views exposure was better than that in the original image exposure, $F(1, 23) = 5.77$, $MSE = 51.042$, $p = .025$, but the difference to the test image exposure did not reach standard levels of significance, $F(1, 23) = 4.18$, $MSE = 35.042$, $p = .052$. In addition, the synthesized views condition had a greater CA score than the original image, $F(1, 23) = 8.115$, $MSE = 41.344$, $p = .014$, and the test image exposure conditions, $F(1, 23) = 5.20$, $MSE = 27.094$, $p = .032$. There was no significant difference between the synthesized views and photographic views conditions, or between the test and original image conditions ($F_s < 1$).

Identification performance in both the photographic views and the synthesized image conditions was better than that of the controls. That is, training exposing either photographic images or synthesized images at multiple yaw rotations facilitated better performance than that of controls exposed to only a limited sample of images. Perhaps most important was the equivalent performance in the synthesized views condition compared to the photographic

**Data analysis**

Two performance measures are reported: The first is identification accuracy, and the second is a confidence-accuracy (CA) measure. The CA score was calculated by multiplying accuracy (negatively scored for incorrect answers so $1 = $ correct and $-1 = $ incorrect) by the confidence score (less 0.5) giving a score between $-6.5$ and $+6.5$ in 13 equal steps. This CA score highlights the fact that a highly confident incorrect answer demonstrates worse performance than low-confidence incorrect answers, while highly confident correct answers represent the best performance.
views condition. This suggests that any important features for recognition of the photographs are being replicated by the photogrammetric software, despite the rather impoverished nature of the synthesized stimuli. It should be noted that the synthesized views exposure may have required less transfer than the photographic views condition, because the synthesized views condition involved exposure to the test angle (although exposure to a generated face at the test angle alone in the test image condition was not sufficient to improve performance—so the benefit of synthesized views does not simply derive from exposure to an image at the test angle). While this is a clear demonstration of the utility of presenting generated images, the conditions did differ in the total number of face images presented. As such, poor performance in the control conditions may have been a product of the relatively small number of exposures compared to the amount given in the multiple view conditions, rather than any advantage gained from multiple angles per se. Experiment 2 examined this issue.

Figure 3. Panel A. Test accuracy as percentage correct (with standard error of the mean, SEM) from Experiment 1. Data are organized by exposure condition (original image, test image, photographic views, and synthesized views). Panel B. Confidence-accuracy (CA) score with SEM. Data are organized as in Panel A.
In Experiment 1 there were several differences between the controls and multiple view conditions that may have contributed to the enhanced performance of providing extra views. First, the total number of exposures rather than the type of exposure may have resulted in better performance of the multiple conditions than that of the control. Secondly, even though the interval between original front-on target face and the presentation of the test array was held constant, the multiple view conditions had a shorter interval between the last face image exposure and the test phase. Therefore, in Experiment 2 (Table 2 summarizes the full design) the photographic views condition from the previous experiment was replaced with a repeated presentation of the original target image (repeated original image). The multiple presentations in a single view enable an assessment of whether the effect observed in the previous experiment, and by Liu et al. (2009), was a product of the amount of exposure given.

### Method

#### Participants

Fifty-one participants, aged between 18 and 24 years, completed the experiment in return for course credit. All participants were recruited from School of Psychology at Cardiff University. None of the participants had participated in Experiment 1. Twenty-four students completed the experiment without the distractor, while 27 participants were tested with the distractor face.

#### Stimuli

All faces were taken from the same set of cropped and computer-generated faces as that used in the previous experiment. Distractor faces were photographs displaying a different gender (i.e., female) that were taken from the same database and were cropped and presented in a fashion identical to that for the other exposed faces.

#### Design and procedure

A within-subjects design gave participants three different exposures (original image, repeated original image, and synthesized views), again using a sequential matching procedure. In this experiment, all exposures began by displaying a front-view face.

### Table 2. Indication of the different training sequences in the six conditions in Experiment 2

<table>
<thead>
<tr>
<th>Condition</th>
<th>Initial exposure</th>
<th>Training sequence</th>
<th>Distractor face</th>
<th>Test arrays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original image</td>
<td>0°</td>
<td>0°, +, +, +, +, +</td>
<td>No</td>
<td>Five 30°L faces</td>
</tr>
<tr>
<td>Repeated original</td>
<td>0°</td>
<td>0°, 0°, 0°, 0°, 0°</td>
<td>No</td>
<td>Five 30°L faces</td>
</tr>
<tr>
<td>Original image</td>
<td>0°</td>
<td>0°, +, +, +, +, +</td>
<td>Yes</td>
<td>Five 30°L faces</td>
</tr>
<tr>
<td>Repeated original</td>
<td>0°</td>
<td>0°, 0°, 0°, 0°, 0°</td>
<td>Yes</td>
<td>Five 30°L faces</td>
</tr>
</tbody>
</table>

Note: All conditions began with the target front-view photo. The training sequence was different for each condition but included fixation crosses “+” and images of the target face at various angles of yaw (indicated by an numerical angle and an indication of whether this was to the right, R, or left, L). The “*” indicates that the image was synthesized using FaceGen rather than being photographic. All distractor conditions ended with the presentation of a female distractor face.

In short, if recognition following repeated exposures is based on image-specific codes (Longmore, Liu, & Young, 2008), or an average representation of that identity (see, Jenkins & Burton, 2008, 2011), then it follows that the synthesized views should facilitate better recognition than the original and repeated original conditions. This is because the multiple angles should allow a better representation of the face, which can generalize when testing recognition at a novel viewpoint.

In Experiment 2 there were several differences between the controls and multiple view conditions that may have contributed to the enhanced performance of providing extra views. First, the total number of exposures rather than the type of exposure may have resulted in better performance of the multiple conditions than that of the control. Secondly, even though the interval between original front-on target face and the presentation of the test array was held constant, the multiple view conditions had a shorter interval between the last face image exposure and the test phase. Therefore, in Experiment 2 (Table 2 summarizes the full design) the photographic views condition from the previous experiment was replaced with a repeated presentation of the original target image (repeated original image). The multiple presentations in a single view enable an assessment of whether the effect observed in the previous experiment, and by Liu et al. (2009), was a product of the amount of exposure given.

### Method

#### Participants

Fifty-one participants, aged between 18 and 24 years, completed the experiment in return for course credit. All participants were recruited from School of Psychology at Cardiff University. None of the participants had participated in Experiment 1. Twenty-four students completed the experiment without the distractor, while 27 participants were tested with the distractor face.

#### Stimuli

All faces were taken from the same set of cropped and computer-generated faces as that used in the previous experiment. Distractor faces were photographs displaying a different gender (i.e., female) that were taken from the same database and were cropped and presented in a fashion identical to that for the other exposed faces.

#### Design and procedure

A within-subjects design gave participants three different exposures (original image, repeated original image, and synthesized views), again using a sequential matching procedure. In this experiment, all exposures began by displaying a front-view face.
photograph of a target individual and ended with a presentation front view of the target. The original image condition was otherwise identical to that of Experiment 1. Similarly, the synthesized views condition gave exposure to multiple computer-generated views as in Experiment 1, although exposures at 30° left were followed by those at 30° right, instead of being intersected with the original image. Presentations in the synthesized condition always ran from 90° left to 90° right, and returned. The repeated original image condition gave the same repeated exposure to the original photograph of a target for the duration of exposure. This was time matched to correspond with the length of presentation time in the synthesized views condition. Exposure times for each stimulus presentation and ISI were the same as those in Experiment 1 (i.e., 2 s per stimulus and a 1-s ISI) meaning that the total exposure time in the no distractor condition was the same as that in the previous experiment. The distractor manipulation used the same exposure sequences as those described above (see Table 2), with the only modification being that the final exposure of a target was followed by a presentation of a novel distractor face at the same angle as the original image (i.e., 0°). These distractor faces were displayed for 2 s followed by a 1-s ISI and then presentation of the test procedure. Different identity distractor faces were used at the end of each exposure condition. Every participant was tested on three different target faces within each experimental condition, and this was counterbalanced such that, across participants, each face was presented in every condition equally often. All other details were the same as those in Experiment 1.

**Results**

Figure 4A displays percentage of correct responses as a factor of condition (original image, repeated original image, synthesized views) and distractor (distractor, no distractor). Performance was generally better in the synthesized and repeated original image conditions than in the original image condition, but there was apparently little effect of distractors. A mixed ANOVA with a within-subject factor of condition and a between-subjects factor of distractor found a main effect of exposure condition on accuracy, \( F(2, 98) = 3.78, \text{MSE} = 0.041, p = .026 \), but no main effect of distractor, \( F(1, 49) = 2.44, \text{MSE} = 0.036, p = .125 \), or interaction between condition and distractor, \( F(2, 98) = 0.59, \text{MSE} = 0.041, p = .557 \). Pairwise analysis suggested that original image exposure produced lower recognition scores than the synthesized views, \( F(1, 49) = 6.896, \text{MSE} = 0.088, p = .011 \), but no other differences were observed [largest \( F(1, 49) = 2.24, \text{MSE} = 0.057, p = .140 \), between original image exposure and repeated original image conditions].

CA scores displayed in Figure 4B indicated a similar pattern of results to that of the percentage correct analysis. There was a main effect of condition, \( F(2, 98) = 5.06, \text{MSE} = 2.571, p = .008 \), but no main effect of distractor, \( F(1, 49) = 1.20, \text{MSE} = 4.106, p = .278 \), or interaction between condition and distractor, \( F(2, 98) = 0.250, \text{MSE} = 2.571, p = .779 \). Pairwise analysis suggested that the original image condition produced lower CA scores than either the synthesized views condition, \( F(1, 49) = 7.00, \text{MSE} = 5.996, p = .011 \), or the repeated original image condition, \( F(1, 49) = 6.912, \text{MSE} = 4.237, p = .011 \), but no difference was observed between synthesized views and repeated original image conditions, \( F < 1 \).

**General Discussion**

The experiments presented here examined the potentially beneficial effects of synthesizing additional training images from a single front-view photograph using a commercially available modeller. This was tested in two experiments using a sequential matching task. The findings of Experiment 1 provide further evidence that extra synthesized face views can aid accurate recognition. Identification following exposure to multiple synthesized images was equivalent to that following multiple photographic views of the target individual, with both conditions superior to exposure to a single front-view photograph. Experiment 2 replicated the advantage for multiple synthesized views over exposure to a front-view photograph alone. However, Experiment 2 also demonstrated that repeating the front-on photograph also improved performance to some extent and resulted in equivalent performance to that for the multiple synthesized views condition.

The results of several studies have indicated that accurate recognition of an unfamiliar face is dependent on the view at which it is presented (e.g., Hill, Schyns, & Akamatsu, 1997; Krouse, 1981; O’Toole, Edelman, & Bülthoff, 1998). Recognition was better when faces were presented in the three-quarter view than in a front-facing view, which in turn was better than in the profile view (e.g., Bruce, Ellis, Gibling, &
Young, 1987; Hill & Bruce, 1996; Liu, 2002). It is thought that the superior 3D information provided by the three-quarter views includes more structural information (Hole & Bourne, 2010). The experiments reported here suggest that synthesizing multiple views—including those akin to a three-quarter view—of a face from a single input can aid recognition beyond that produced by the original photo of an unfamiliar face alone (see also, Liu et al., 2009). This is consistent with the finding that virtual busts, created using photogrammetric software, can facilitate recognition close to the level of recognition produced by real photographic stimuli presented during training (Bailenson et al., 2004).

Taken together with the present study, these results indicate the potential for utilizing artificially generated face images to overcome the viewpoint dependence associated with unfamiliar faces. According to Liu et al. (2009), the ability to synthesize an angle close to that of a test view can help bridge the gap between original photo and test image. This is consistent with evidence from the object and face recognition literature that indicates that it is easier to generalize based on exposures to multiple views rather than a single view (e.g., Edelman & Bülthoff, 1992; Hill et al., 1997). The fact that recognition can be supported on the basis of low-cost synthesized images suggests that even relatively impoverished

Figure 4. Panel A. Test accuracy as percentage correct (with standard error of the mean, SEM) from Experiment 2. Data are organized by exposure condition (original image, repeated original image, and synthesized views) and are presented as a function of distractor type (distractor, no-distractor). Panel B. Confidence-accuracy (CA) score with SEM. Data are organized as in Panel A.
stimuli may be of some assistance in situations where only limited images are available.

While it is important that recognition can be supported in any way on the basis of photogrammetric images, it did not reliably exceed repeated exposure to a single photograph in Experiment 2. There are two general possible explanations for this pattern of results. First, the SI FaceGen software provides impoverished stimuli and thus does not support perfect transfer because it loses information when making the synthesized images. As such, even relatively impoverished stimuli may be of some assistance, but the level of support derived from such images may be limited. This possibility implies that a superior face generation algorithm would support better performance than repeated exposure to a single image. Secondly, that there is a fundamental limit on the overlap between the information common to different views of a face—and that the human visual system is just as good at extracting this common information as is any computer program following multiple exposures to an individual. Obviously some combination of both of these possibilities might well be in operation. In addition, the potential practical benefit of using generated faces will depend on the degree to which superior generation software can overcome the limitations of the system used in the current experiments.

To explore these ideas in more detail—consider the differences between the synthesized and photographic images used here (see Figure 2), in particular the contrast in surface pigmentation between images. If, as Longmore et al. (2008) suggest, familiar faces that have been seen many times allow for extraction of structural codes, then the period of familiarization should have allowed extraction of some of these codes. This is supported by the evidence that viewing time is, at least in part, related to improvements in recognition accuracy (e.g., Dennett, McKone, Edwards, & Susilo, 2012; Memon, Hope, & Bull, 2003). These structural codes are thought to include information similar to that of object recognition and include three-dimensional shape as well as surface pigmentation (Marr & Nishihara, 1978). Thus, repeated exposure to a single photo in the current studies may support the acquisition of surface pigmentation, but it is unlikely that much of the shape information is conveyed. In contrast, comparison across synthesized images at different angles (spanning the three-quarter view is thought to convey more structural information than front-on or profile views) may allow the extraction of three-dimensional shape information. However, these synthesized images may lack the surface pigmentation detail that would be afforded by real photos or a better modeller. While nothing in the current experiments, or those of Liu et al., 2009 directly addresses this possibility, the fact that the current software produces images that are clearly less than photographic quality and lack accurate textural information would certainly be consistent with the idea that superior modelling software could afford even better support for face learning across different viewing angles.

In summary, the current experiments reinforce prior demonstrations that providing multiple views of an unfamiliar face will improve subsequent identification performance relative to exposure to a single viewpoint. Most importantly, the results also suggest that synthesized images—generated from a single original photograph—can also support improved recognition in a similar fashion. The degree of improvement produced by these generated images may have been limited by the fact that they were somewhat impoverished (especially in terms of detailed textural information), implying that superior modelling software should support larger beneficial effects. Regardless, the fact that even the current set of stimuli facilitate processing of unfamiliar faces strongly suggests that artificially synthesizing multiple views of a face from limited photographic input could be a very valuable technique in overcoming the fallibility of processing of unfamiliar faces in the context of limited image sets.

**Funding**

The research reported here was undertaken at Cardiff University and was funded by a PhD studentship to Scott Jones from the Engineering and Physical Sciences Research Council, United Kingdom. Information on data supporting the results presented in this article, including how to access those data, can be found in the Cardiff University research-data catalogue at http://dx.doi.org/10.17035/d.2016.0008381942.

**ORCID**

Scott P. Jones http://orcid.org/0000-0001-5516-4385

Dominic M. Dwyer http://orcid.org/0000-0001-8069-5508

Michael B. Lewis http://orcid.org/0000-0002-5735-5318

**References**


