Law and Human Behavior, Vol. 31, No. 2, April 2007 (© 2007)
DOI: 10.1007/s10979-006-9007-2

Catching the Bad Guy: Morphing Composite Faces Helps

Lisa E. Hasel[1,2,3] **and Gary L.Wells**[2]

Published online: 12 October 2006

Composite faces built by eyewitnesses commonly are poor likenesses of the target face. When there are multiple witnesses, however, an opportunity exists to morph the composites. Morphs were rated as more similar to the target face than were the mean ratings of the individual composites. However, as hypothesized, the morph also came to resemble non-target faces more than the individual composites did (a prototype effect). This prototype effect was so strong that the morphs resembled non-targets more than the individual composites resembled the targets. In addition, morphing composites produced an attractiveness bias, which made the morphing of composites less effective for less attractive targets. Even when the prototype effect and the attractiveness bias were controlled for, however, a true morph-superiority effect continued to exist.

KEY WORDS: eyewitness; composite; morph; face; multiple witnesses.

> A convenience store was robbed and three eyewitnesses individually worked with police to produce facial composites of the culprit. Each produced a very different face, so the police released all three composites on a wanted poster. Police later received a message from a nearby small town police department that read: "Have arrested two of the suspects and are hot on the trail of the third."
>
> (story that circulates among U. S. law enforcement agencies)

Although undoubtedly apocryphal, this illustrates a common problem for crime investigators: Eyewitnesses often give differing descriptions of the same culprit. This is part of a more general problem with using eyewitnesses to identify culprits (Wells & Olson, 2003).

When lineups cannot be conducted because no specific suspect exists, composite faces are commonly used by law enforcement. Composite faces are generated by having witnesses work with sketch artists or using a composite-production system, such as the Identi-Kit, Photo-Fit, CD-Fit, Mac a Mug, or FACES. Unfortunately, composite faces generally are poor likenesses of the intended target faces (e.g.,

[1]The author won the Psi Chi/APS Albert Bandura Graduate Research Award in 2005–2006 for this study.
[2]Iowa State University, Ames, Iowa.
[3]To whom correspondence should be addressed at Psychology Department, Iowa State University, West 112 Lagomarcino Hall, Ames, Iowa 50011; e-mail: lhasel@iastate.edu.

Bruce, Ness, Hancock, Newman, & Rarity, 2002; Ellis, Davies, & Shephard, 1978; Kovera, Penrod, Pappas, & Thill, 1997). Early research suggested that the poor likenesses produced by composite systems might be attributable to the systems themselves (e.g., Ellis et al., 1978). However, much of the difficulty lies in the fact that composites require people to recall isolated facial features whereas memorial representations of faces tend to be holistic (Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Wells & Hryciw, 1984). Accordingly, all composite systems tend to perform poorly (Frowd et al., 2005).

If there were multiple eyewitnesses, would it help to average (morph) the composites? We have reproduced a result from the current research in Fig. 1. Notice how the individual composites do not look as similar as the morph does to Photographed Face #1. Hence, it would appear in Fig. 1 that morphing these composites produced a better likeness of the target face than did the individual composites. However, the issue turns out to be more complex than it first appears. In fact, Photographed Face #1 in Fig. 1 is not the target, but rather a filler. The actual target is

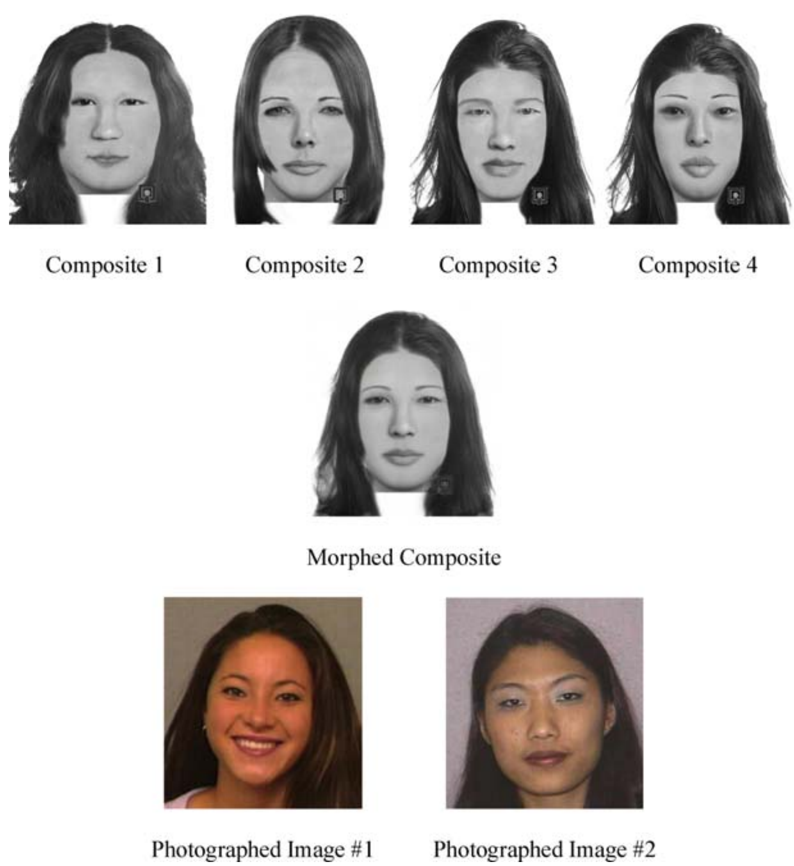

Fig. 1. Four individual composites, a morph of the four individual composites, and the comparison photographs.

Photographed Face #2. Notice, however, that the morph also better resembles the actual target (Photographed Face #2) than do the individual composites. This result illustrates the operation of two spurious influences that we hypothesize occur from the morphing of composite faces, namely a morph-attractiveness effect and a prototype effect. In order to test the hypothesis that morphing composites produces a better result than the individual composites, we designed a study that tests for the existence of (and then controls for) these two spurious influences.

The proposition that morphing individual composites could produce a better product than the individual composites was introduced by Bruce et al. (2002), who reported that morphing four individual composites resulted in a better likeness of the target than did the mean of the individual composites. Their first experiment used publicity photographs of four actors (e.g., George Clooney) and four models, and their second experiment used photographs of four psychology department staff members.

Although Bruce et al. (2002) did not articulate the underlying theory, the implicit assumption is that each facial composite has error, the errors are largely uncorrelated, and these errors are randomly distributed around some true value representing the target face. Hence, we might expect morphed composites to be superior to the individual composites. Before accepting the conclusion reached by Bruce et al., however, controls needed to be implemented over two hypothesized spurious effects related to morphing faces that could each account for part of their results. Specifically, we hypothesized a morph-attractiveness effect and a prototype effect as two potential artifacts of morphing composites that may have played into the findings by Bruce et al.

Spurious Contributions of the Morph-Attractiveness Effect

There are many studies showing that morphing the faces of different people produces a new face that is more attractive than the mean attractiveness of the individual faces that went into the morph (e.g., Langlois & Roggman's, 1990; Langlois, Roggman, & Musselman, 1994; Rhodes & Tremewan, 1996; Rhodes et al., 2001; Valentine, Darling, & Donnelly, 2004). Could the morph-attractiveness effect have played a role in the results reported by Bruce et al. (2002)? Recall that the four targets that Bruce et al. used in Experiment 1 were publicity photographs of famous actors and photographs from a casting agency catalog of models. On average, we would expect publicity photographs of actors and models to be attractive images of those peoples' faces. This raises the possibility that the apparent increases in similarity from morphing the composites were due, at least in part, to the morph-attractiveness effect. The issue of attractiveness is less clear in Bruce et al.'s Experiment 2, in which the four target faces were photographs of psychology department staff members. We note, however, that the effect sizes for the morph-superiority effect were larger for Experiment 1 (using attractive photographs of actors and models) than for Experiment 2 (using photographs of psychology staff). In any case, the current experiment used target faces that varied in attractiveness so we could measure and analyze the role of attractiveness in the morph-superiority effect.

There are no data on the attractiveness of morphed computer composite faces and it is not a foregone conclusion that morphing composite faces increases attractiveness. Rhodes, Roberts, & Simmons (1999) suggested that morphed faces are more attractive than individual faces because morphing increases symmetry and symmetrical faces are considered to be more attractive than nonsymmetrical faces (see Kalick, Zebrowitz, Langlois, & Johnson, 1998; Thornhill & Gangestad, 1999). On the other hand, Valentine et al. (2004) found the morph-attractiveness effect even for profiles and concluded that the morph-attractiveness effect "is independent of any effect of symmetry" (p. 482). Because composite faces used in the current research were highly symmetric, our research provides another test of whether the morph-attractiveness effect is attributable merely to increases in symmetry.

Spurious Contributions of a Prototype Effect

A second possible spurious contribution of morphing composite faces is that the morphed face might simply be more prototypical of the general category of faces from which the target is drawn, thereby producing an increase in similarity to all or most faces from that category. By prototype, we mean the average or central tendency of the members of a category (Smith & Medin, 1981). In the case of faces, categories might be represented in terms of general description categories involving gender, race, age, and hair. Consider the category of White males in their 20s with short, dark hair. Presumably there is a central tendency of such faces, perhaps represented by averaging those faces, and the result of averaging could be called the prototypical White male in his 20s with short, dark hair.

It is not our purpose to press a particular view of the concept of prototype or category. We simply note that any claim for the morph-superiority effect must control for any increase in similarity to relevant comparison faces. Consider a situation in which a blond-haired, White male in his 50s committed a robbery that was seen by four eyewitnesses. Suppose that each eyewitness created a distinctively different composite face of a White male in his 50s with blond hair. A morph of the four composites might better resemble the robber's face than the individual composites do, but the morph might also better resemble most White males in their 50s with blond hair than do the individual composites because the morph has moved the image closer to the prototype of the general category.

If there is a prototype effect from morphing, then the greater similarity of the morph to the target than the individual composites to the target has two possible components: a genuine morph-superiority component and a prototype component. Our study was designed to test for the prototype component and remove its contribution in order to see if a true morph-superiority effect remained. Our design used 16 target faces, four from each of four different categories. Every face was used equally often as a target for which four composites were created by participants. A morph of the four composites per target was created. A new sample of participants then rated the similarity of each composite and morph to their respective targets as well as to the other three non-target (filler) faces within that category. A prototype effect can then be estimated by calculating how much more similar the morph is to the fillers than the individual composites are to the fillers. This prototype effect can then be subtracted from how much more similar the morph is to the target than the individual composites are to the target. It is only this residual (i.e., the interaction contrast) that can represent the true morph-superiority effect because any increase in similarity that the morph produces to the target must be greater than increases in similarity to non-targets from that target's category.

METHOD

Participants

All participants in this study were undergraduate students at a large Midwestern university who received partial course credit for their participation. Sixteen people participated in the composite generation phase, 72 people participated in the first attractiveness evaluation phase, and 60 people participated in the similarity evaluation and second attractiveness evaluation phase of this study.

Materials

Target Faces

We selected 16 target faces from a large database of facial photographs. The 16 photographs used in this study were of 19–23-year olds that fell into four different categories: Four Asian females with long black hair, four Caucasian females with short brown hair, four Caucasian males with short brown hair, and four Caucasian males with medium-length black hair. Within each category, one face was selected that was more attractive and one face was selected that was less attractive than the other faces in the group to ensure that there were differences in attractiveness within a category.

Composite Software

The composite program was FACES: The Ultimate Composite Picture (Cote, 1998), which is described in detail elsewhere (Wells, Charman, & Olson, 2005). The program provides participants with 17 different categories of facial features (e.g., hair, chin, ears, eyes). Within each category there are between 33 (goatees) and 593 (noses) instances of a feature from which composite-makers can choose. Once a feature is selected, it can be moved or sized according to the directions of the composite-maker. The FACES composite system is very popular, with over 150,000 copies sold worldwide. In the United States, FACES is second only to the Identikit in its usage by law enforcement (McQuiston & Malpass, 2000). Recent comparisons between EvoFIT, FACES, PROfit, and E-FIT in terms of their success in creating likenesses of the intended faces revealed very similar performances of these systems (Frowd et al., 2005).

Morphing Software

The morphing program was MorphMan 4.0 (2003, Moscow, Russia) by STOIK Imaging. Using MorphMan, two composites of the same face were morphed together by aligning critical points (i.e., eyes, nose, lips, ears, chin, face shape, hair edge, and wrinkles) on one composite with corresponding points on the second composite. MorphMan averages the grey levels located in a fixed matrix for the critical points, thus creating an average of the two images. A transition image was created that depicted 50% of the first composite image and 50% of the second composite image. This process was repeated for the two other composite images of the same face. The two morphed images were then morphed together to create the final morph, which contained 25% of each of the four composites that went into it.

Procedure

Composite Generation and Morphing Phase

Sixteen participants in this phase created four composites apiece. Participants were told upon arrival to the study that they were going to be utilizing facial composite software called FACES (Cote, 1998). Training consisted of a demonstration by the experimenter of what categories of features were available, how to search for and select features within a category, and how to move and resize features once they were selected. The experimenter ensured that the participants felt comfortable using the software before showing the participants the first face that they would attempt to recreate.

For each of the 16 target faces, four separate participants saw the face for 1 min. After the face was removed, each participant worked with the FACES (Cote, 1998) program to build a composite to his or her satisfaction. Although no specific time limit was set, no participant spent more than 25 min completing any one composite. Each participant created four composites, one from each of the four categories of faces, yielding 64 individual composites, four for each of the 16 faces. After building each composite, the composite builders rated their own composites for how similar they thought their composite was to the face they had viewed on a scale from 1 (*no similarity*) to 10 (*very similar)*. The four composites for each face were then morphed together by an independent experimenter. This resulted in a total of 80 face images: 64 individual composites and 16 morphed composites (see all composite, morph, and target images at http://www.psychology.iastate.edu/faculty/gwells/morphstimuli.pdf).

Attractiveness Evaluation Phase One

A sample of 72 participants rated the attractiveness of every face image on a scale from 1 (*very unattractive*) to 10 (*very attractive*). Participants first viewed a sheet of paper with all 16 of the actual faces on it and were asked to rate the attractiveness of each face. They then viewed a sheet of paper with the 16 composites and four morphs of the faces within a category on it. Morphed composites were always on the bottom row of this page. This was repeated for each category of faces, and the order in which participants saw the composite-and-morph sheets was counterbalanced. These participants' attractiveness ratings were used only to compare the overall attractiveness of the three face image categories (photos of target faces, composites, morphs).

Similarity Evaluation and Attractiveness Evaluation Phase Two

A separate sample of 60 participants rated the similarity of each target's composites and the morphed product of those composites to both the target and to the other three non-target faces from that target's category. Each participant received a packet with one page in it for each target face. Every page had the four faces in the target face's category across the top and the four composites and the morph for that target face along the side (see how the similarity rating sheets were composed at http://www.psychology.iastate.edu/faculty/gwells/morphstimuli.pdf). The order of the photographed faces was counterbalanced within and across participants, as was the order of the composites and morph. The order that the face arrays were shown to participants was randomized, with the caveat that no two face arrays were presented sequentially that contained faces of the same gender.

Each rater made 20 similarity judgments for each of the 16 targets, for a total of 320 similarity judgments from each of the 60 raters. Consider, for example, the 20 ratings made by a given participant for one of the four Caucasian males with short brown hair. Participants rated the similarity of each of the four composites of this target to both the target photo and to the photos of the other three Caucasian males with short brown hair (16 ratings). They also rated the similarity of the morph of the four composites to both the target and to the other three Caucasian males with short brown hair (four ratings). This was then repeated across all 16 targets. Similarity ratings were made on a scale from 0 (*totally different*) to 100 (*identical*).

After completing the similarity evaluation, participants were asked to review their ratings. They looked at each row (i.e., the ratings for one composite or morph for each of the faces in a face array) and were asked to circle the highest similarity rating within each row. If there was a tie for the highest rating, then they were asked to circle the number of the comparison that they thought looked *most* like the computer-generated face. This provided us with a target-preference rate. Finally, the same 60 participants rated the attractiveness of every target face on a scale from 1 (*very unattractive*) to 10 (*very attractive*).

RESULTS

Inferential statistics on the similarity data were based on standardized scores at the level of each participant to reduce variance in how the 60 participants used the similarity scale. All descriptive statistics are reported using the actual scale values. We used the mean similarity rating of the four composites for each the 16 targets for purposes of comparing them to the similarity rating for each of the 16 morphs. In addition, we used the mean similarity rating of the three fillers for each of the 16 face arrays for purposes of comparing them to the similarity scores for the 16 targets. We protected the overall α level using the Bonferroni correction for the 11 pairwise comparisons of similarity ratings. Therefore, differences were not considered significant unless they were significant at the .0045 or less level.

Overall Attractiveness Analyses

Analyses of the attractiveness ratings indicated that the photographic faces were more attractive than the morphed images, $t(71) = 3.75$, $p = .001$, $d = .94$. The morphed images, in turn, were more attractive than the mean of the individual ratings of the composites, $t(71) = 8.20$, $p < .001$, $d = 2.05$.

Overall Similarity Analyses

Our first similarity analysis collapsed the ratings across the 16 targets. An analysis of variance was then conducted on the 2 (image: morph versus composite) $\times$ 2 (array member: target versus filler) design. Both main effects and the interaction were significant. Morphs produced significantly higher similarity scores than did individual composites, $F(1, 59) = 482.10$, $p < .001$, Cohen's $f = 2.85$ and similarity to targets was higher than was similarity to fillers, $F(1, 59) = 421.57$, $p < .001$, Cohen's $f = 2.67$. Importantly, an interaction was obtained, indicating that the increase in similarity resulting from morphing was significantly greater for the targets than for the fillers $F(1, 59) = 103.44$, $p < .001$, Cohen's $f = 1.32$. It is this interaction, depicted in Fig. 2, which represents the morph-superiority effect.

Despite evidence of a genuine morph-superiority effect, contrast analyses conducted within the 2×2 design indicated a spurious elevation in similarity resulting from morphing because the morphs showed higher similarity to the fillers than

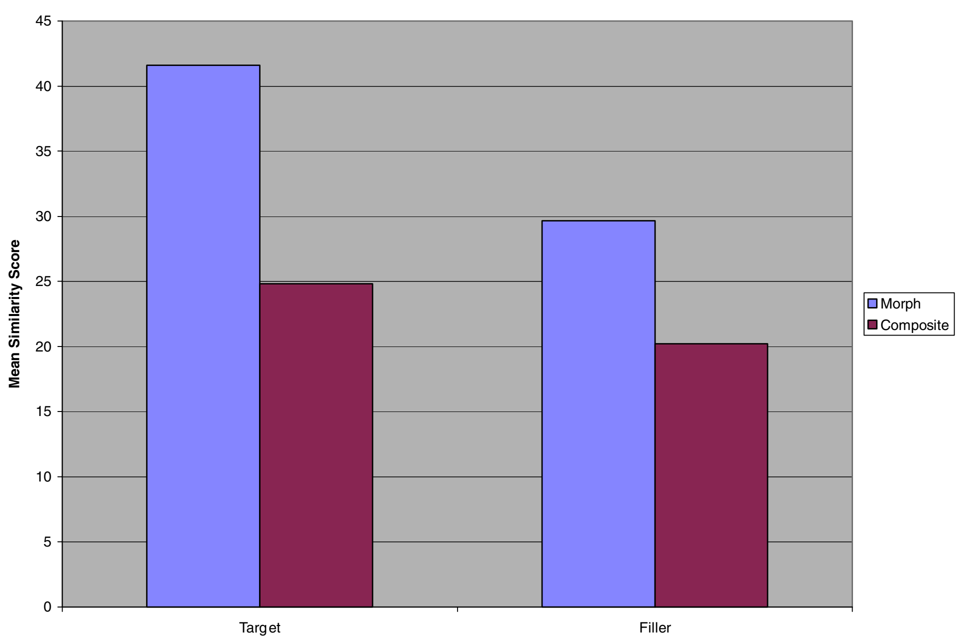

Fig. 2. Overall similarity ratings for morphs to targets, individual composites to targets, morphs to fillers, and individual composites to fillers.

did the individual composites, $t(59) = 7.26$, $p < .001$, $d = 94$. It is this contrast (the two right-hand bars in Fig. 2) that captures the notion of a prototype effect resulting from morphing. This prototype effect was large ($d = .94$) and was over half the size of the morphs-versus-individual-composites effect for the targets, $t(59) = 12.95$, $p < .001$, $d = 1.67$. In fact, the prototype effect was so large that the similarity of the morph to the non-target fillers was significantly greater than the similarity of the individual composites to the actual targets, $t(59) = 3.40$, $p = 001$, $d = 44$.

Similarity Analyses by Attractiveness

Our second analysis involved a quartile blocking on the relative attractiveness ratings of the 16 targets (the four most attractive targets, next four most attractive, next four most attractive, four least attractive within each face array). Using this blocking on attractiveness, we conducted a 4 (attractiveness: first, second, third, fourth quartiles) × 2 (image: morph versus individual composite) × 2 (array member: filler versus target) ANOVA to see if the three-way interaction was significant. The three-way interaction was significant, indicating that attractiveness moderated the morph-superiority effect, $F(3) = 9.61$, $p < .001$, Cohen's $f = .40$. The morph-superiority pattern of means held true for all four levels of target attractiveness (the two left-hand bars show a greater difference than the two right-hand bars), $F(1, 59) = 52.188$, $p < .001$; $F(1, 59) = 50.10$, $p < .001$; $F(1, 59) = 4.653$, $p = .04$; and $F(1, 59) = 7.97$, $p = .006$, from most to least attractive target quartiles, respectively. However, as can be seen in Fig. 3, a clear pattern emerged in which the magnitude of the two-way interaction (hence, the strength of the morph-superiority effect) was stronger for the top two quartiles of target attractiveness (Cohen's f's = .94,.92) than the bottom two quartiles (Cohen's f's = .28, and .38).

The prototype effect is estimated by the magnitude of the difference between the similarity scores for the morphs versus the individual composites for the fillers (the two right-hand bars for each graph in Fig. 3). Across all four levels of target attractiveness, contrasts indicated that this simple effect was significant, $t(59) = 4.94$, $p < .001$; $t(59) = 6.46$, $p < .001$; $t(59) = 5.54$, $p < .001$; and $t(59) = 5.15$, $p < .001$, for the most attractive to least attractive target quartiles, respectively. These effect sizes were not systematically related to target attractiveness, ds = .64, .83, .71, and .66, respectively. Hence, as would be expected, the prototype effect itself was not systematically affected by target attractiveness. Instead, the advantage of the morph over the individual composites in similarity to the target decreased as target attractiveness decreased (the two left-hand bars of each graph in Fig. 3), $t(59) = 11.70$, $p < .001$, $d = .1.51$; $t(59) = 13.06$, $p < .001$, $d = 1.69$; $t(59) = 7.25$, $p < .001$, $d = .94$; $t(59) = 7.42$, $p < .001$, $d = .95$, for the most to least attractive target quartiles, respectively.

As further evidence for the role of target attractiveness, we calculated a morph-superiority interaction term at the level of each participant for each target using the standardized scores. We then regressed each participant's standardized attractiveness rating for each target on each participant's own morph-superiority interaction term. The mean correlation was $r = .15$, $t(59) = 5.14$, $p < .001$, $d = .66$, with a mean

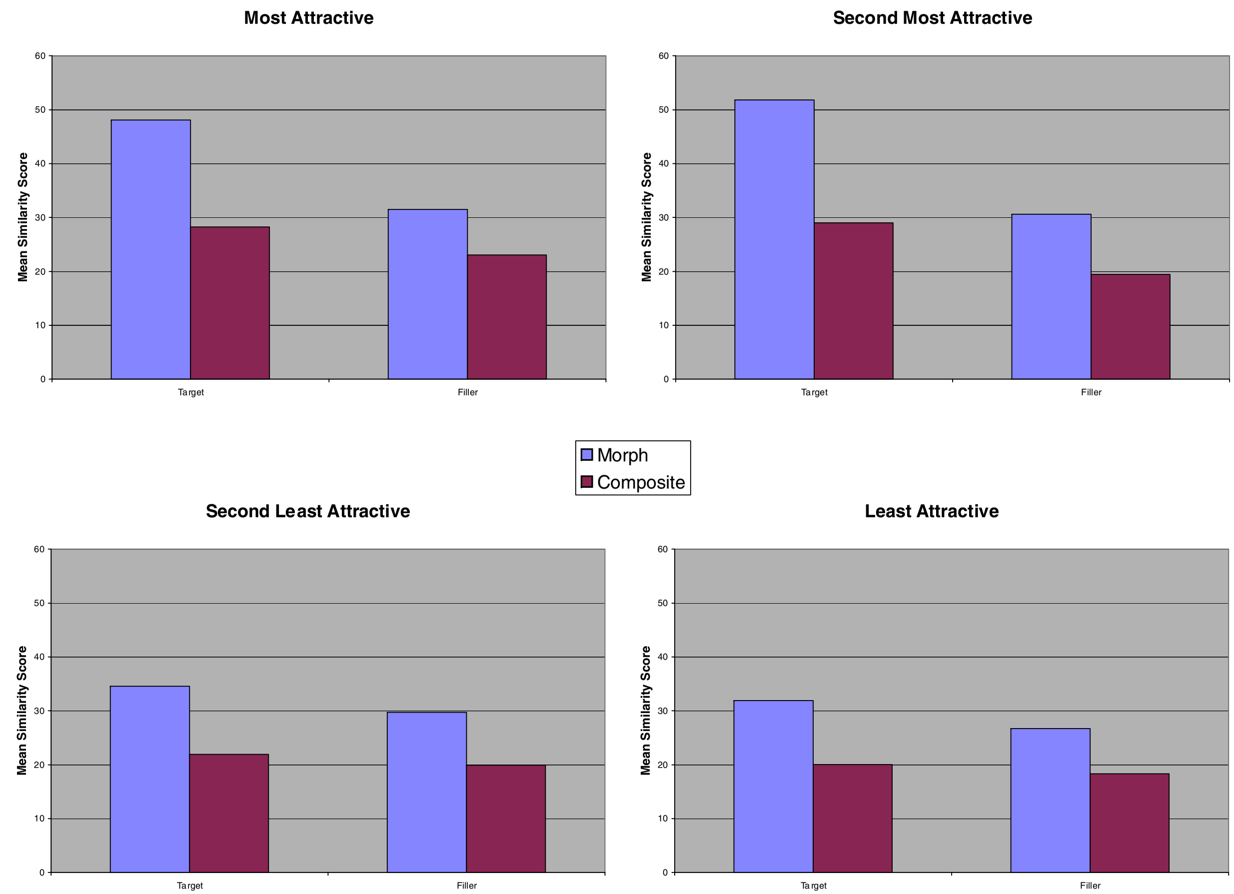

Fig. 3. Similarity ratings for morphs to targets, individual composites to targets, morphs to fillers, and individual composites to fillers for each quartile of target attractiveness.

slope of .20, $t(59) = 4.66$, $p < .001$, $d = .60$. This is further evidence indicating that the magnitude of the morph-superiority effect depends partly on target attractiveness. Specifically, every standard deviation in attractiveness has a .20 standard deviation impact on the morph-superiority effect. Important for our purposes was the mean intercept of .39, which was significantly greater than zero, $t(59) = 9.88$, $p < .001$, $d = 1.28$, indicating that the morph-superiority effect remained significant even at the lowest level of target attractiveness (lowest attractiveness rating on 1–10 scale = 3.18, highest = 8.17, $SD = 1.36$). This indicates that a face would have to be almost two standard deviations in attractiveness below the least attractive target that we used to reach a point where there is no morph-superiority effect. Two standard deviations below the least attractive target would be less than 1.0 on the 1–10 scale, which, of course, is not possible. This suggests that the morph-superiority effect should remain above zero for all faces even though the effect size would become miniscule for the least attractive faces.

Target-Preference Rates

Participants selected one person from the face array to whom they thought each composite or morph looked the most similar. If the participant selected the target, then this was scored as target-preference, but if the participant selected one of the fillers, then this was scored as a "miss." Note that the target-preference scores, unlike the similarity scores, automatically control for any prototype effects because each face array contains both the target and fillers. Hence, any genuine morph advantage for the target would have to exceed the prototype effect on the fillers in order to elevate target-preference rates.

Collapsed across targets, morphs produced higher target-preference rates (47.7%) than did individual composites (35.2%), $t(59) = 7.45$, $p < .001$, $d = .96$ (see Fig. 4). Furthermore, the morphs produced higher target-preference rates than did the individual composites for three of the four quartiles of target attractiveness, $t(59) = 7.36$, $p < .001$; $t(59) = 4.51$, $p < .001$; $t(59) = .96$, $p = .34$; $t(59) = 2.43$, $p = .02$, respectively. As with the similarity analyses, the advantage of the morph over the individual composites diminished as target attractiveness decreased (ds = .95, .58, .12, and .31, respectively). Finally, both the morphs and the individual composites produced target-preference rates significantly greater than the chance rate of 25.0%, $t(59) = 13.95$, $p < .001$, $d = 1.80$ and $t(59) = 12.15$, $p < .001$, $d = 1.56$, respectively.

Morphs Versus Best Individual Composites

Our final analyses involved comparisons between morphs and the best individual composites (rather than the means of the individual composites). The question here is whether there are individual composites that are better than the morph. We defined the best individual composites in two different ways. First, we identified which one of the four individual composites for each of the 16 faces yielded the highest target-preference rate. We will call this method of defining the best individual composites as the "post-hoc method" because the "best" can only be

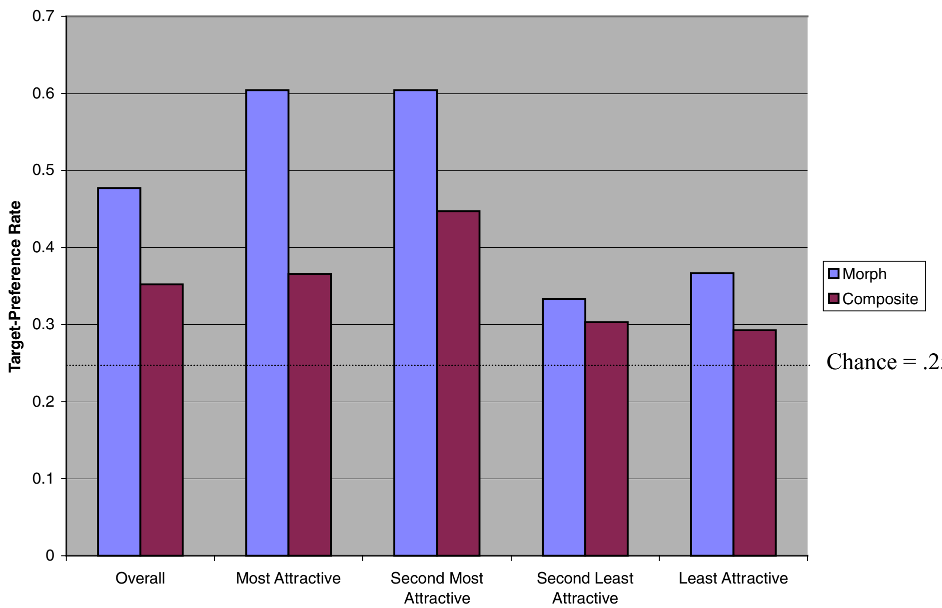

Fig. 4. Target-preference rates for morphs and for individual composites; overall and for each quartile of target attractiveness.

determined by knowing which face is the actual target. The post-hoc method cannot be used in an actual case because it presumes that the target face is already known. The other method we used for defining "best" was to use the self-ratings given by the composite builders. Recall that composite builders rated their own composite for how similar they thought their composite was to their memory of the face that they were trying to build. For each target face, the individual composite that was rated the highest by the composite builder was defined as the best composite. We will call this method for defining the best individual composite the "a-priori method" because it does not require one to know which face is the target in order to define the best composite. The a-priori method can be used in actual cases to sort among individual composites because it relies only on the self-reports of the composite builders.

Using the post-hoc method, the target-preference rate for the best individual composites was .49, which was not significantly different from the .48 target-preference rate for the morphs, $t(15)=.26$, $p=.79$, $d=.07$. Using the a-priori method, the target-preference rate for the best individual composite was .36, which was significantly lower than the .48 target-preference rate for the morphs, $t(15)=.2.72$, $p=.018$, $d=.68$. These results indicate that the morph does not do better (or worse) than the best individual composite if somehow we knew which individual composite was the best. When using an a-priori method for deciding which individual composite is the best (i.e., the self-ratings of the composite builders), the morph is better.

DISCUSSION

Morphing individual composite faces produces an image that better resembles the target face than does the mean rating of the individual composites. After controlling for any spurious contributions, the effect appears to be real. Nevertheless, some of this effect is an illusion in the sense that the morphed product also shows greater resemblance to non-target faces than does the mean rating of the individual composites. This supports our hypothesis that morphing composite faces produces a prototype effect and that some of the apparent increase in similarity is not distinctive to the target face itself. As can be seen in Fig. 2, more than one-half of the apparent benefit of morphing individual composites is actually a prototype effect.

We also found that morphing individual computer composites produces a face that is more attractive than are the individual composites that went into the making of the morph. Because the composite faces yielded by the FACES program were almost perfectly symmetric (except for hair), this result lends further support to Valentine et al.'s (2004) claim that the morph-attractiveness effect for faces is not attributable to mere increases in the symmetry of the face. More importantly for our purposes, however, is that the strong moderating role of target attractiveness meant that the morph-superiority effect (after controlling for the prototype effect) was strong when the target face was attractive (mean Cohen's $f = .93$ for the top two quartiles of attractiveness) but much weaker when the target was less attractive (mean Cohen's $f = .33$ for the lower two quartiles of attractiveness). The regression analysis on the interaction terms indicated a statistically significant advantage of morphing composite faces even after removing the influence of attractiveness. Accordingly, our results support Bruce et al.'s (2002) contention that there is a morph superiority effect for composite faces.

Clearly, some individual composites are better than others. The best individual composites, identified using the post-hoc method, produced target-preference rates comparable to the morphed composites (48.9% and 47.7%, respectively). The a-priori method (composite builders ratings of their own composite), however, generally failed to identify the best individual composites. The target-preference rate for the highest self-rated composites was 35.6% whereas the overall target-preference rate for individual composites was 35.2%. This suggests that there is little utility to asking eyewitnesses to evaluate their own composites. It suggests further that producing morphs that weight the individual composites according to the self-ratings of the composite builders would not produce a better morph, a finding consistent with what Bruce et al. (2002) reported. Perhaps there are better questions that can be asked of witnesses regarding their own composites, but we found that asking them a straightforward question regarding how similar they thought their composite was to the actual face was not effective in identifying the best composites.

There are, of course, other composite systems and other morphing programs that might produce different results. However, we note that we used both a different composite program and a different morphing program than did Bruce et al. (2002) and yet our results largely replicate theirs. In general, we might expect that the better the individual composites the less advantage the morph would show over the individual composites. We note, however, that the individual composites produced in the current study appear to be slightly better than what is usually obtained in composite studies. Kovera et al. (1997), for example, reviewed 10 studies and noted that the average hit rate was 28%. The composites in the current work showed an average preference rate for the target over similar-looking fillers of 35%. Had we found that the morphs were not superior to the individual composites, we might worry that the morphing program that we used was simply not sophisticated enough to reveal a morph-superiority effect. However, the program that we used, which simply averages the grey levels located in a fixed matrix for the critical points, was sophisticated enough to reveal the morph-superiority effect. More sophisticated programs, such as one that averages the spatial locations and shapes of anatomical features (see Pittenger, 1991), might produce even better results.

Despite evidence for a true morph-superiority effect after controlling for spurious effects of morphing, our findings raise some questions about the purpose of composites. If the purpose of a composite image is to reduce the pool of possible suspects, the morphing of individual composites is likely to backfire. The prototype effect itself is quite strong ($d = .94$), which means that morphing individual composites leads to an increase in similarity to other (nontarget) faces. As a result, it is likely that morphing individual composites will result in more (rather than fewer) innocent people appearing similar to the final product. If, on the other hand, the purpose of a composite is to help ensure that the actual culprit is among those who resemble the composite, then morphing is likely to help serve that purpose.

Our results present an encouraging picture for the prospect of producing better composites when there are multiple eyewitnesses. Furthermore, the morphing of composites is quick, simple, and inexpensive. However, as with any technique, there is a potential to use it in ways that could be misleading. In particular, we are concerned about someone treating a morphed composite's similarity to a defendant as a form of evidence against the defendant. For example, we can imagine a prosecutor showing a jury that individual composites of the perpetrator do not look much like the defendant but the morphed composite looks a great deal like the defendant: "Yes, the individual composites were not very good, but look what happens when we average them!" This could appear to be a compelling type of evidence used against a criminal defendant, but it could be an illusion. We have shown how this result is expected to occur even when the suspect is not the culprit because of the prototype effect. At the very least, in order to consider the evidence to be probative of guilt, the increase in similarity resulting from morphing composites would have to be shown to be greater for the suspect than for non-suspects who also happen to fit that general physical description.

REFERENCES

Bruce, V., Ness, H., Hancock, P. J. B., Newman, C., & Rarity, J. (2002). Four heads are better than one: Combining face composites yields improvements in face likeness. *Journal of Applied Psychology*, *87,* 894–902.

Cote, P. (1998). *FACES: The ultimate composite picture*(Version 3.0). [Computer software]. IQBiometrix: Quebec, 467. Canada: Saint Hubert.

Ellis, H. D., Davies, G. M., & Shepherd, J. W. (1978). Remembering pictures of real and unreal faces: Some practical and theoretical considerations. *British Journal of Psychology*, *69,* 467–474.

Frowd, C. D., Carson, D., Ness, H., McQuiston-Surrett, D., Richardson, J., Baldwin, H., et al. (2005). Contemporary composite techniques: The impact of a forensically-relevant target delay. *Legal and Criminological Psychology*, *10,* 61–83.

Kalick, S. M., Zebrowitz, L. A., Langlois, J. H., & Johnson, R. M. (1998). Does human facial attractiveness honestly advertise health? Longitudinal data on an evolutionary question. *Psychological Science*, *9,* 8–13.

Kovera, M. B., Penrod, S. D., Pappas, C., & Thill, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology*, *82,* 235–246.

Langlois, J. H., & Roggman, L. A. (1990). Attractive faces are only average. *Psychological Science*, *1,* 115–121.

Langlois, J. H., Roggman, L. A., & Musselman, L. (1994). What is average and not average about attractive faces? *Psychological Science*, *5,* 214–220.

McQuiston, D., & Malpass, R. S. (2000). *Use of facial composite systems in U.S. law enforcement agencies*. Poster presented at the American Psychology—Law Society, New Orleans, LA.

Pittenger, J. B. (1991). On the difficulty of averaging faces: comments on Langlois and Roggman. *Psychological Science*, *2,* 351–353.

Rhodes, G., Roberts, J., & Simmons, L. W. (1999). Reflections on symmetry and attractiveness. *Psychology, Evolution, and Gender*, *1,* 279–295.

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and attractive faces. *Psychological Science*, *7,* 105–110.

Rhodes, G., Yoshikawa, S., Clark, A., Lee, K., McKay, R., & Akamatsu, S. (2001). Attractiveness of facial averageness and symmetry in non-Western cultures: in search of biologically based standards of beauty. *Perception*, *30,* 611–625.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A,* 225–246.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, *25,* 583–592.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Science*, *3,* 452–460.

Valentine, T., Darling, S., & Donnelley, M. (2004). Why are average faces attractive? The effect of view and averageness on the attractiveness of female faces. *Psychonomic Bulletin and Review*, *11,* 482–487.

Wells, G. L., Charman, S. D., & Olson, E. A. (2005). Building face composites can harm lineup identification performance. *Journal of experimental psychology: Applied*, *11,* 147–156.

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory and Cognition*, *12,* 338–344.

Wells, G. L., & Olson, E. (2003). Eyewitness identification. *Annual Review of Psychology*, *54,* 277–295.