Building Face Composites Can Harm Lineup Identification Performance

Gary L. Wells Steve D. Charman

Iowa State University

Elizabeth A. Olson

University of Wisconsin - Whitewater

Author contact:

Gary L. Wells

e-mail: glwells@iastate.edu

phone: 515/294-6033

fax: 515/294-6424

Abstract

Face composite programs permit eyewitnesses to build likenesses of a target faces by selecting facial features and combining them into an intact face. Research shows these composites to generally be poor likenesses of the target face. Two experiments tested the proposition that this composite-building process could harm the builder's memory for the face. Experiment 1 (n = 150) used 50 different faces and found that the building of a composite reduced the chances that the person could later identify the original face from a lineup compared to no composite control conditions or to yoked composite-exposure control conditions. Experiment 2 (n = 200) found that this effect generalized to a simulated-crime video, but mistaken identifications from target-absent lineups were not inflated by composite building.

Building Face Composites Can Harm Lineup Identification Performance

Eyewitness identification from live and photographic lineups is a staple type of evidence used against criminal suspects. Over the last 25-30 years, cognitive and social psychologists have conducted research questioning the accuracy of eyewitness identification evidence, examining the critical variables, and devising methods to improve eyewitness identification accuracy (e.g., see Cutler & Penrod, 1995, Wells, 1993). By the late 1990s, the use of forensic DNA testing began to corroborate the research conclusions of psychologists by showing that mistaken eyewitness identification was the primary cause of jury convictions of innocent people that were later overturned using forensic DNA tests (Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). The U.S. justice system has now started to incorporate better methods of eyewitness identification based on psychological research (see Wells, Malpass, Lindsay, Fisher, & Turtle, 2000).

When criminal investigators have both a crime suspect and an eyewitness to the crime, investigators place the suspect (or a photo of the suspect) in a lineup of persons (or photos of persons) for the purpose of seeing if the eyewitness can identify the suspect as the culprit in question. Variables affecting the chances that the witness will accurately identify the culprit, identify an innocent person, or make no identification at all are highly complex. These variables include witness characteristics (e.g., Hosch & Platz, 1984), viewing conditions at the time of the crime (e.g., Lindsay, Wells, & Rumpel, 1987), disguises (e.g, Cutler, Penrod, & Martens, 1987), witness and culprit race (see Meissner & Brigham, 2001), retention interval (e.g., Krafka & Penrod, 1985), pre-lineup instructions given to the witness (see Steblay, Dysert, Fulero, & Lindsay, 1997), and characteristics of the lineup fillers that are used (e.g., Wells, Rydel, & Seelau, 1993), and whether or not the lineup contains the actual culprit (see Wells & Olson, 2002), among others.

Often, crime investigators have an eyewitness to a crime but have no particular suspect. In such cases, investigators will sometimes resort to the use of techniques to build a facial composite of the culprit. Police sketch artists, for example, were used in as many as 10% of law enforcement agencies in the mid-1980s (MacDonald, 1984), but these have tended to be replaced by other facial composite production systems. The Identi-kit and the Photofit Kit, for example, use transparencies of facial features that are superimposed over each other to create facial likenesses. Hundreds of hairstyles and eyes, dozens of mouths, noses, and chins are available for the person to select when building the face. In more recent years, computerized versions of facial composite systems have been developed to run on PCs. One of the most common of the computerized systems in use in U.S. law enforcement agencies today is the FACES program, which is what we used in the current research. All of these facial composite systems have in common the fact that the person must select individual facial features (e,g,. hair, nose, eyes, chin, mouth, facial hair) and combine them to yield an emergent image of a face. This "particularistic" (feature based) approach used by composite systems is quite different from the presumed "holistic" manner in which people are presumed to perceive and store faces in memory (see Wells & Hryciw, 1984; Tanaka & Farah, 1993; Tanaka, & Sengco, 1997).

Although the meaning of holistic is not well understood at a theoretical level, it seems reasonable to conclude that faces are not represented in memory in a manner that permits people to build facial composites that represent reasonable likenesses of the target face that they intended to build. Research has consistently shown that various facial composite systems yield rather poor matches to the original faces (e.g., Ellis, Davies, & Shepherd, 1978; Ellis, Shepherd, & Davies, 1975; Christie & Ellis, 1981; Gibling & Bennet, 1984; Green & Geiselman, 1989). Using a computerized facial composite system, for example, Kovera, Penrod, Pappas, & Thill (1997) found that composites of individuals who were well known to the composite builder were no better than chance in directing naïve participants to the correct person in a five person lineup that included the target individual[1].

If composites tend to be poor likenesses of the target face that they were intended to depict, might this have implications for the witness's memory for the face and their prospects for being able to identify that face later? There are, of course, several possibilities. First, it is possible that building a composite face has no implications for the witness's memory; the witness's memory for the original face remains intact and the composite face does not interfere or compete with that memory. Second, it is possible that building a composite face creates a second memory; there are now two memories of the face, the original memory and the composite memory, and they compete in any later memory task. Third, it is possible that the original memory is blended or averaged with the original memory, yielding a new face that has some characteristics of the original face and some of the composite face. Fourth, it is possible that the original memory of the face is replaced with the composite face, yielding only one face in memory, namely that of the composite. There are also hybrid possibilities that involve more than one of these processes. For instance, the witness might end up with three faces in memory; the original face, the composite face, and a blend of the two. These ideas of two memories, replacement of the original memory, blended memories, and so on are actually quite complex and have not been fully resolved in related literatures that involve suggested memories (e.g., see Loftus, Schooler, & Wagenaar, 1985; McCloskey & Zaragoza, 1985). Accordingly, the current experiments will not be able to resolve the precise process. However, we can establish some things. First, we can establish whether or not building a composite harms the ability of a person to identify the original face from a lineup. Second, we can establish whether any such effect is due to a shift in confidence (or a decision-criterion shift) or whether it is the result of a reduced familiarity of the original face. Third, we can establish whether any such effect is due to mere exposure to a composite face or whether the composite-building process itself causes the effect. Finally, we can establish how the effect of building a composite operates in lineups in which the target face is not present, whether it leads to increased correct rejections or actually increases the chances that an innocent person will be identified.

It could be said that building a composite is similar to giving a verbal description of a face, except that a composite is visual and is more concrete and specific. Indeed, the primary reason that crime investigators use composites is because eyewitnesses' verbal descriptions of culprits are usually too few in number and too vague (see Sporer, 1996; Wells, Rydell, & Seelau, 1993) to lead to specific suspects. Note, however, that both verbal descriptions and composites share the characteristic of being feature-based productions for what is presumed to be holistic memorial representations of the face. Accordingly, research on verbal overshadowing could shed some light on what to expect from composite productions. The verbal overshadowing effect is generally described as a negative effect of verbalizing a face on later attempts to recognize that face. The verbal overshadowing effect, although usually small, has been replicated extensively in the literature (e.g., Dodson, Johnson, & Schooler, 1997; Fallshore & Schooler, 1995; Ryan & Schooler, 1998; see meta-analysis by Meissner & Brigham, 2001). Recent experiments, however, suggest that the verbal overshadowing effect is the result of a change in decision criterion rather than an alteration of the underlying memory trace (Clare & Lewandowsky, 2004). Clare & Lewandowsky reached this decision-criterion conclusion based on two primary observations in their experiments. First, the verbalization reduced choosing rates for both target-present and for target-absent lineups. Second, when forced to make a choice from a target-present lineup, there was no effect of the verbalization manipulation. Hence, in the current work, we wanted to be sure that we could distinguish between a criterion shift and some type of changed memory that might result from building a composite. In Experiment 1, therefore, we permitted a "not present" response even though the target face was present but then forced non-choosers to make a choice. Also, in Experiment 2, we included target-absent lineups to see if building a composite increased the rate of non-choosing.

We also wanted to be able to distinguish between the composite-building process itself and mere exposure to a composite face of the target. When a witness builds a composite, the witness must sort among a large number of facial features and settle on those seem to best represent his or her memory of the target face. The witness can place these features together into a face and then try out various features, replacing them at will, until she or he is satisfied that the product is the best that can be achieved. In the end, a whole face is examined by the witness. We included conditions in the current experiments in which a second participant was exposed to the resulting composite produced by each participant. These participants (hereafter called “yoked composite-exposure participants”) were informed that this was a composite of the target that was built by another participant. It is possible that any detrimental effects in the composite-building conditions on later lineup identifications are the mere result of having viewed a final composite of the target, not the process of having built the composite.

We believe that the questions that we are asking in this research require stimulus sampling for all the reasons described recently by Wells and Windschitl (1999). Specifically, we need some assurances that the faces that are used in this research are not unique in some way that make them especially likely to show or not show any effects. Hence, we sampled 50 different faces for use in Experiment 1, we used each face equally often in each condition, and we built 50 different six-person lineups around each face. This methodology prevented us from using realistic crime scenarios in Experiment 1 because this would have required us to develop 50 video or live crimes around each of the 50 targets. However, once we had established the generality of the effect across a broad sample of faces in Experiment 1, we used a video crime in Experiment 2.

We predicted that building face composites would have two different effects. First, we predicted that participants who built a face composite would show a strong criterion shift as evidenced by an increased reluctance to attempt an identification. This effect is similar to Clare & Lewandowsky’s (2004) claim about what happens with verbal overshadowing. In effect, it is a reduced confidence on the part of the eyewitness that she or he knows what the target looks like. However, we also expected that building a composite would actually damage the memory of the witness. Unlike verbal descriptions, the composite-building process requires visual processing of isolated competing facial features that could blend with or replace features of the target. Furthermore, the composite process yields a final visual product of a specific intact face. Hence, we expected that the composite-building participants would show a diminished capacity to be able to identify the target when forced to make an identification. It was less clear what to predict for the yoked composite-exposure conditions. Unlike the composite-building conditions, participants in the yoked composite-exposure conditions are not exposed to isolated facial features and, although they view a final visual product of an intact face, that face is from another participant, thereby permitting them to possibly separate that image from their own memory. Hence, although we expected that the yoked composite-exposure participants would perhaps experience some reduction in their confidence in their ability to identify the target, it might not actually harm their ability to identify the target if forced to make an identification.

Experiment 1

Method

Participants and design. One hundred and fifty undergraduate students participated in exchange for extra credit in their psychology courses at Iowa State University. Participants were randomly assigned to one of three conditions; a control condition, a composite-building condition, or a yoked composite-exposure condition. Within each condition, participants were randomly assigned to view one of 50 target faces. A caveat to random assignment required that the composite-building condition always have N+1 completed composites compared to the yoked composite-exposure condition so that there was always a composite from a composite-building participant to show to a subsequent yoked composite-exposure participant. Sessions involved either one or two participants per session, each being placed in a separate room with a computer.

Materials. The 50 target photos were of 29 white males and 21 white females ranging in age from 18 years to 27 years. For each of the 50 target photos, five filler photos were selected who matched each target photo on the characteristics of gender, race, approximate age appearance, hair color, hair length, and any facial hair. In some cases, filler photos were repeated if they fit appropriately in a lineup for more than one target. Because a given participant only viewed one lineup, this did not result in a participant seeing the same photo twice. A total of 212 filler photos were used for the 50 six-person lineups. All photos had similar backgrounds. Photos were high-quality color photographs from a straight-on pose showing the neck to the top of the head. No clothing clues were visible. Each photo was presented on a color computer screen with the image measuring 8 inches by 6.4 inches. The composite program used was FACES: The Ultimate Composite Picture (1998) by Interquest Incorporated. FACES includes 361 hair selections, 63 head shapes, 42 forehead lines, 410 sets of eyebrows, 514 sets of eyes, 593 noses, 561 sets of lips, 416 jaw shapes, 145 moustaches, 152 beards, 33 goatees, 127 sets of eyeglasses, 70 eye lines, 147 smile lines, 50 mouth lines, and 40 chin lines. Within each feature category, a selection button permits the user to view only subsets of the feature that meet a particular description. For instance, eyes are subdivided into the subsets narrow, deep set, overhanging lids, heavy lids, average blue or green, almond shaped blue or green, average brown, almond shaped brown, and bulging. Noses are subdivided into the subsets of narrow, average with round base, average with broad base, average pointed, hooked nostrils not showing, hooked nostrils showing, slightly flared nostrils, very flared nostrils, round (bulbous), average large, wide base with nostrils showing, and side base with nostrils not showing. In addition, controls permit the features to be moved up or down, closer or farther apart, and made to be larger or smaller. Features are displayed on the right side of the computer screen. When clicked with a mouse, the feature then appears on the left side of the screen in a position appropriate to that feature on a face. So, for example, clicking any head shape, hair, eyebrows, eyes, nose, lips, and chin yields a very realistic face. Clicking a different set of eyes then instantly replaces the eyes on that face with the new eyes. Footprints keep track of features attempted so that the operator can return easily to a previous feature. Details can then be added such as the mouth lines, smile lines, facial hair, and so on. Distances between features and feature sizes on the face can be manipulated on the intact face on the left side of the screen. The product can then be saved electronically. The FACES program is very user-friendly.

Procedure, day 1. Participants signed up for an experiment on "perceptions of people" but otherwise were told nothing about the purpose of the study before signing up. Participants understood that they would have to return for a second session 48 hours later. On arrival at the lab, participants were told that they would be asked to view a face on a computer screen and make various judgments about the face. Participants were then escorted individually to a cubicle that had a computer screen, keyboard, and mouse. The experimenter then gave each participant a sheet of paper with 10 question items. The ten items asked the participant to rate a face on the traits attractive, intelligent, warm, aggressive, kind, happy, foolish, humorous, studious, and likeable. An 11-point scale for each item ranged from 0 (not all all) to 10 (very). The experimenter then presented the target face on the computer screen. Participants were given three minutes alone to complete their ratings while the face remained in view. The experimenter then returned and eliminated the target's image from the screen. Participants in all conditions then were asked to provide a verbal description of the target face on a blank sheet of paper. After writing their descriptions, control participants were dismissed with instructions to return 48 hours later for additional questions. After writing their descriptions, composite-building condition participants were instructed on the use of the FACES program and attempted to build the target face. Instruction took about eight minutes. The participant was then left alone to use the composite program. The average participant took about 16 minutes to create a composite. The experimenter then saved the composite product on the computer and the participant was dismissed at that point with instructions to return 48 hours later for additional questions. After writing their descriptions, participants in the yoked composite-exposure condition were told (correctly) that another participant had viewed the same face that they had just viewed and then had used a composite program to try to build that face. They were then shown the composite from their yoked participant and informed that this was the result of that participant's attempt. Participants in the yoked composite-exposure condition were dismissed at that point with instructions to return 48 hours later for additional questions.

Procedure, day 2. Only two participants failed to show 48 hours later for their second appointment and those two are totally excluded from this report. On their return, participants in all conditions were treated identically. Participants first filled out a questionnaire that took approximately 12 minutes and is totally unrelated to the current research. Then, participants were escorted back to their original cubicle. It was explained that we were interested in their ability to identify the photo of the person they had viewed and rated on various traits 48 hours earlier from a lineup of six photos. The experimenter strongly emphasized that the photo of the person that they had rated 48 hours earlier might or might not be among the six that they were going to view. “Hence,” the experimenter explained “the correct answer might be ‘none of the above.’” In fact, the target was always in the lineup and appeared in a position determined randomly for each participant. The lineup arrays were displayed as two rows of three persons each, all six displayed simultaneously. After the identification decision was recorded, any participants who made no identification were then asked to go ahead and make an identification to wit “If you had to identify one of these photos as the one that you rated 48 hours ago, which photo would you pick?” Participants then rated their confidence that their identification was correct on a scale from 0% (not at all confident) to 100% (positive).

Results

Table 1 reports the percentages of participants who identified the target, a filler, or made no identification in each of the three conditions. The fourth column reports the percentages of participants who identified the target after the non-identifying participants were forced to identify someone. A series of Chi Square tests were performed to locate statistically reliable differences between conditions in the rates of identification of the target, identification of fillers, no identification decisions, and identifications of the target when forced to choose. The results showed very strong evidence that both building a composite and being exposed to the composite of another participant reduced identifications of the target relative to the control, χ^2 (1, N = 100) = 54.96, $p < .05$, χ^2 (1, N = 100) = 17.36, $p < .05$, respectively. In addition, the building of a composite reduced identifications of the target relative to being exposed to a yoked composite, χ^2 (1, N = 100) = 14.66, $p < .05$. Although there were no differences between the control condition and the yoked composite-exposure condition on rates of filler identification, building a composite inflated filler identifications relative to both of those conditions χ^2 (1, N = 100) = 9.76, $p < .05$, χ^2 (1, N = 100) = 9.76, $p < .05$, respectively. Both the building of a composite and exposure to the composite of another participant increased the rate of no identifications, χ^2 (1, N = 100) = 24.47, $p < .05$, χ^2 (1, N = 100) = 19.05, $p < .05$, respectively. However, the rate of no identifications was not significantly greater for those who built a composite than it was for those who were merely exposed to the composite of another participant χ^2 (1, N = 100) = 1.01, ns. When participants were forced to choose someone from the lineup, there was no significant difference in the rates of identifying the target for the control participants versus those who were exposed to the composite of another participant, χ^2 (1, N = 100) = 3.41, ns. However, the participants who built a composite continued to show a diminished rate of being able to identify the target relative to both the control participants and those who were exposed to the composite of another participant χ^2 (1, N = 100) = 43.46, $p < .05$, χ^2 (1, N = 100) = 27.44, $p < .05$, respectively.

Confidence results. Participants' mean confidence in identifications was significantly higher for the control condition (M = 78.3%) than for either the composite building condition (M = 45.6) or the yoked composite-exposure conditions (M=56.3%), ts (99) = 3.34 and 2.21, respectively, $p_s < .05$, whereas the latter two conditions did not differ, t (99) = 1.12, ns. The point-biserial confidence-accuracy correlation was quite high for the control condition and for the composite building condition, $r_{pb} = .59$ and .54, $p_s < .05$, respectively, but quite low for the yoked composite-exposure condition, $r_{pb} = .12$, ns.

Discussion

The results of this experiment provide evidence that building a face composite diminishes the prospects that a person will later be able to identify that face from a six-person lineup. Compared to a control condition, both those who built a composite face and those who merely observed the composite face that was built by another participant exhibited a five-fold increase in the tendency to make no identification attempt from the lineup at all. However, when forced to choose someone from the lineup, only the participants who built a composite face exhibited a significant decreased ability to identify the target face. Hence, although mere exposure to the face composite product of another participant reduced participant’s confidence that they could identify the target and increased their reluctance to make an identification, mere exposure, unlike actually building a composite face, did not appear to affect the ability of participants to accurately identify the target.

The confidence-accuracy correlations revealed a particularly interesting pattern of results. Participants in the control condition and participants in the composite-building condition exhibited a strong confidence-accuracy relation, but those in the yoked composite-exposure condition did not. We propose that this pattern can be explained rather easily. Those who were exposed to the composite of another participant tended to lose confidence in their ability to identify the target from the lineup and yet, when they were forced to make a choice, were in fact quite accurate. They seemed to not be aware that they in fact knew the correct answer. Those who built a composite also lost confidence that they could identify the target from the lineup and, unlike their yoked composite-exposure controls, were in fact quite inaccurate when forced to make a choice. Hence, the participants who built a composite seemed quite aware that they in fact did not know the correct answer.

Our goal in this experiment was to determine whether building a face composite creates some damage to the person's memory of the target face. We believe that we have established that it does cause such damage, at least under the conditions in which we operationalized this experiment. Our yoked composite-exposure control condition helps rule out interpretations that are based merely on exposure to a completed composite face. Nevertheless, we do not know the precise mechanisms that are producing the effect. We strongly suspect that the principal damage results from the process of having to break the face down into individual facial features in order to perform the composite building task. There is rather strong consensus in the extant literature on face processing that faces induce holistic, Gestalt-like processing and representation that reduces the accessibility of information about individual facial features (see review by Mauer, Le Grand, & Mondloch, 2002). Although there is debate about precisely what is meant by holistic, configural processing of faces, it seems to us that forcing people to try to recreate a holistic memory by recognizing and assembling individual parts is an unnatural act that can reconfigure the memory representation.

There are, of course, other levels of explanation for the results that we could not test in this experiment. For instance, the performance difference between the yoked composite-exposure control condition (which did not show memory damage) and the composite build condition (which did show damage) could have been due to the differential “investment” that they had in the composite. Those who were merely exposed to a composite of another participant had no personal investment in that product, no particular reason to believe that it was a serious attempt by the other participant, and so on. On the other hand, yoked composite-exposure did have a very strong effect on participants’ responses, cutting the hit rate in half under conditions in which they were given a choice of making no identification and profoundly lowering their confidence when they were forced to make a choice. Hence, mere exposure to a composite face did have strong effects and so participants clearly did not dismiss the composite as something that was uninformative about the identity of the target.

We have learned from related literatures on misinformation effects that it is almost impossible to determine the fate of the original memory, such as whether the original memory has been altered by post-event information or whether a second memory has been formed that competes with the original memory at the time of test (e.g., see Loftus & Loftus, 1980; Loftus, Schooler, & Wagenaar, 1985; McCloskey & Zaragoza, 1985). It is worth noting, however, that our experiment used a version of the so-called “modified test” proposed by McCloskey and Zaragoza that helps rule out the two-memory hypothesis. McCloskey and Zaragoza noted that misinformation experiments typically include the misinformation item as a test alternative. Such tests could simply prove that the participants have two memories, one for the original item and one for the misinformation item. The modified test, on the other hand, is one in which the original item remains an option among filler items but the misinformation item is not one of the options. The McCloskey and Zaragoza modified test was used in this experiment because it included the original item (target-present lineups) but not the misinformation item (e.g., the composite or a face that was designed to resemble the composite). If participants in the composite-build conditions had two memories (one of the original face and one of the composite), then they still should have been able to identify the original face. The fact that they could not identify the original face when the composite face was not an option suggests that the original memory was actually impaired.

An applied perspective was one of the primary driving factors in this research and the applied implications of this experiment are fairly clear. The decision by crime investigators to build a composite face is usually based on the issue of whether or not there is a suspect in the case. If there is a suspect, or a defined set of possible suspects, then investigators will usually follow those leads, check alibis, and perhaps conduct a lineup for the eyewitness. Composites tend to be reserved for cases in which there are few or no leads and the idea is that the composite might help flesh out possible suspects. The results of this experiment, however, suggest that the decision to develop a composite comes at a potential cost. The cost can occur later when a suspect emerges in the investigation and the eyewitness is then shown a lineup containing that suspect. Given the usual cautionary instructions (the actual culprit might or might not be present in this lineup), participants who built a composite in this experiment were less than 1/8 as likely to identify the actual target than were those who did not build a composite. If forced to choose someone, they remained less than 1/3 as likely to be able to pick out the target. There might be other costs and benefits to the decision to have an eyewitness build a composite of a culprit, but the profound diminishment the eyewitness's later ability to pick the culprit out of a lineup is a serious consideration that must be brought into the decision process.

Experiment 2

The applied implications of this work beg for a further test of the effect of composite building, namely what happens if the witness views a lineup that does not contain the actual target? Extensive writings and experiments in the eyewitness identification literature explain the importance of including target-absent lineups in eyewitness identification research designs (e.g., see recent treatment in Wells and Olson, 2002). Because a proper lineup contains only one suspect (who might or might not be the culprit), eyewitness identification researchers do not count identifications of fillers in a target-present lineup as mistaken identifications of innocent suspects. A mistaken identification of an innocent suspect, therefore, can only occur when the culprit is not in the lineup (see Wells & Turtle, 1986). Furthermore, identification patterns with target-absent lineups are not fully predictable from identification patterns with target-present lineups. A primary purpose of Experiment 2, therefore, was to find out whether the building of a composite would increase mistaken identification rates in target-absent lineups. The results of Experiment 1 suggest that this is likely to happen because building composites increased the rate of filler identifications in target-present lineups. Hence, we predicted that building a composite would reduce accurate identification rates for target-present lineups (as in Experiment 1) but also increase mistaken identifications for target-absent lineups. Although we expected composite building to increase mistaken identifications for target-absent lineups, we also expected the increase to be limited by the tendency of composite building to make participants less confident in their ability to make an identification.

Another reason for conducting the second experiment was to increase the ecological validity of the materials. Experiment 1 used still photos for encoding and still photos for the lineup. The target photo for the lineups in Experiment 1 was the same photo that was used for encoding. This circumstance creates a great advantage in overall performance because the participant need only recognize the specific photo of the target rather than a different photo of that person. This is perhaps the reason that the overall performance of the control condition participants was so high (94% able to pick out the target photo when forced to make a pick). In actual criminal cases involving eyewitnesses, however, the witnessed event is more dynamic and the photo of the target used in a target-present lineup is not an exact duplication of what the witness viewed. In an actual criminal case, the lineup photo of the target in a target-present lineup is the same person who committing the criminal act, but not an exact duplicate of what the witness originally saw (e.g., not the same expression on the face, not the same pose, photo not taken on the same day as the crime, and so on). Accordingly, Experiment 2 used a video crime and the photo of the culprit was taken on a different day. We did not include a yoked composite-exposure condition in Experiment 2.

Method

Participants and design. Two hundred undergraduate students participated in exchange for extra credit in their psychology courses at Iowa State University. Participants were randomly assigned to one of four conditions of a 2 (Target-present or Target-absent lineup) X 2 (Composite build versus Control) factorial design. Sessions involved either one or two participants per session, each being placed in a separate room with a computer.

<u>Materials</u>. The witnessed event was a video lasting 65 sec showed a man on a rooftop placing an object down an air shaft. His face was in full view for 21 sec. The man was approximately 20-22 years old with short, dark hair, no facial hair or glasses, and no other distinguishing characteristics. Hereafter, he is called the rooftop bomber or target. Seven high-quality color photographs were used to create the photo lineups. One photo was of the rooftop bomber and the other six photos were other males who were approximately 20-22 years old with short, dark hair, no facial hair or glasses, and no other distinguishing characteristics. The target-absent lineup included all six of the non-target photos and were always in the same order. Target-present lineups were created by replacing each member of the target-absent lineup with the target eight times for positions 3-6 and nine times for positions 1-2 in each target-absent condition . [Because there were 50 participants and six potential positions for the target, two positions were occupied one extra time. However, this was equally true for both the control condition and the composite condition.] This replacement scheme means that there was no a-priori innocent suspect for the target-absent lineups; hence, the unbiased estimate of the chances that an innocent suspect would be identified is 1/6 of the total rate of identifications of someone from the target-absent lineup.

<u>Procedure, day 1</u>. Participants signed up for an experiment on "perceptions of people" but otherwise were told nothing about the purpose of the study before signing up. Participants understood that they would have to return for a second session 48 hours later. On arrival at the lab, participants were told that they would be asked to watch a short video and to pay close attention as they would be asked questions about it later. Participants were then escorted individually to a cubicle that had a computer screen,

keyboard, and mouse. The video was then played for the participant on the computer screen. After the video ended, the experimenter returned and informed participants that the person on the roof in the video had planted a bomb down the air shaft of the building. Participants in all conditions then were asked to provide a verbal description of the rooftop bomber's face on a blank sheet of paper. After writing their descriptions, control participants were dismissed with instructions to return 48 hours later for additional questions. After writing their descriptions, composite-building condition participants were instructed on the use of the FACES program and followed the same procedure that was used in Experiment 1.

Procedure, day 2. Only three participants failed to show 48 hours later for their second appointment and those three are totally excluded from this report. On their return, it was explained to all participants that we were interested in their ability to identify the bomber on the roof that they had viewed 48 hours earlier. The experimenter strongly emphasized that the photo of the person that they had rated 48 hours earlier might or might not be among the six that they were going to view. "Hence," the experimenter explained "the correct answer might be 'none of the above.'" The lineup arrays were displayed as two rows of three persons each, all six displayed simultaneously. After the identification decision was recorded, any participants who made no identification were then asked to go ahead and make an identification: "If you had to identify one of these photos as the one that you rated 48 hours ago, which photo would you pick?" Participants then rated their confidence that their identification was correct on a scale from 0% (not at all confident) to 100% (positive).

Results

Target-present lineups. Table 2 reports the percentages of participants who identified the target, a filler, or made no identification for the target-present lineups for both the control and composite build conditions. The fourth column reports the percentages of participants who identified the target after the non-identifying participants were forced to identify someone. A series of Chi Square tests were performed to locate statistically reliable differences between the two conditions in the rates of identification of the target, identification of fillers, no identification decisions, and identifications of the target when forced to choose. The results replicated those of Experiment 1. Building a composite reduced identifications of the target relative to the control, χ^2 (1, N = 100) = 18.54, $p < .05$. Compared to the control condition, building a composite also increased the rate of filler identifications, χ^2 (1, N = 100) = 6.06, $p < .05$, increased the rate of no-identification decisions, χ^2 (1, N = 100) = 13.04, $p < .05$, and reduced the rate of identifications of the target when participants were forced to make an identification, χ^2 (1, N = 100) = 32.67, $p < .05$.

Target-absent lineups. Table 3 reports the percentages of participants who made correct rejections and who identified someone from target-absent lineups. There were no effects of composite building on the rates of correct rejections, χ^2 (1, N = 100) = 0.51, ns. No Chi Square test was conducted on rates of identification from the target-absent lineups because this is statistically identical to the test of correct rejections (i.e., 100% - correct rejection percent = identification percent).

Confidence results. As in Experiment 1, participants' mean confidence in identifications was significantly higher for the control condition (*M* = 72.8%) than for the composite building condition (*M* = 53.5%) within the target-present conditions, t (99) = 4.33, $p < .05$. Unlike Experiment 1, however, the point-biserial confidence-accuracy correlation was low for both the control condition (r_{pb} = .17, ns.) and for the composite building condition (r_{pb} = .12, ns.). There was no significant confidence difference between the control (M = 46.7%) and composite build conditions (M = 43.1%) within the target-absent conditions t (99) = 1.03, ns. Because confidence was measured only after forcing an identification (i.e., confidence in non-identifications was not measured), no confidence-accuracy correlation could be calculated in the target-absent conditions.

Discussion

Using different materials, we replicated the deleterious effects that composite building has on rates of identifying the target from a target-present lineup. Participants who built a composite were less likely to attempt an identification and when forced to make a selection were far less likely to make an accurate selection. This increases our confidence that the effects observed in Experiment 1 might generalize to actual witnesses who observe dynamic events and attempt identifications from photo-lineups that include the culprit, but do not use an exact replica of the face as it was seen originally.

Interestingly, we found no significant effects of composite building on the rates of correct rejections (and hence on mistaken identification rates) for the target-absent lineups. Although we had expected that there would be more mistaken identifications from the composite building conditions based on their greater rate of filler identifications in Experiment 1, in retrospect these results make sense. Even if composite building alters memory for the original face, there is no reason to think the altered memory systematically resembles the members of the target-absent lineup any more than the original face resembles members of the target-absent lineup. Accordingly, there is no reason to believe that control condition participants (whose memory of the target is not altered) would find a match to the original face at a rate lower than would those whose memory was altered as long as the target face is not in the array. In this sense, participants in the control and composite build conditions are largely equal when the target is not in the lineup.

General Discussion

We began this work with the goal of seeing if having a person build a composite of a face has a negative effect on the person's memory for the face. Mindful of the need for stimulus sampling, we used 50 separate faces in Experiment 1 in order to make sure that our conclusions could be generalized across a broad sample of faces. In addition, we wanted to separate any effects of mere exposure to the composite face products from the process of actually being the one who built the composite face, so we yoked each participant who built a composite to another participant who was then exposed to the composite product. Furthermore, we wanted to separate any effects of a criterion shift (specifically, an increased reluctance to attempt an identification) from an actual change to or confusion of the original memory. Hence, participants were first given the option of making no identification, but then all participants were forced to make a lineup choice. The results clearly supported the hypothesis that building a composite harmed memory for the original face. When forced to make a choice from a lineup that included the original target face, accurate identifications of the face were reduced dramatically for those who built a composite face. Those who were exposed to a composite produced by another participant were also strongly affected, but the effect was limited to a criterion shift in which these yoked participants were more reluctant to attempt an identification, but when forced to make a choice showed no impaired ability to accurately identify the original target.

Our second experiment showed that the results can be generalized to more dynamic witnessing events in which the lineup photos at test are not duplicates of the exact stimulus face that was encoded, a condition that applies to virtually all actual lineup situations. In addition, our second experiment shows that, other things being equal, the building of a composite does not inflate mistaken identification rates for target-absent lineups. The latter observation, however, needs to be associated with a very important caveat. It is clear from Table 3 that we found no significant increase in mistaken identification rates for target-absent lineups as a function of composites. Nevertheless, this research did not test a particular set of conditions that might be especially likely to inflate mistaken identifications in actual cases. Specifically, the primary purpose of having eyewitnesses build composites in actual cases is to narrow the pool of potential suspects and perhaps receive "tips" regarding the identity of the culprit. This is why, for instance, composites are often published in newspapers or broadcast on television news programs. If an innocent person then becomes a suspect because of resemblance to the composite built by the eyewitness, and thus ends up being placed in a lineup, the innocent suspect could very well match the "new" memory of the eyewitness and be at high risk of mistaken identification. [This concern applies equally regardless of whether we assume the composite changes the witness's memory or simply creates a second memory.] We did not test this potentially dangerous path. In other words, our choice of members for the target-absent lineups were not influenced by the composites themselves. It was not our purpose to test this hypothesis, but we think it is a serious limitation to our conclusion that building composites did not inflate the chances of mistaken identifications from target-absent lineups.

The legal system has taken a great deal of direction from the eyewitness identification literature in recent years and the impact of this lab-based psychological science has been remarkable (see Wells, Malpass, Fisher, Turtle, & Fulero, 2000). Interestingly, however, almost all of psychology's research-based recommendations to the legal system have emphasized law enforcement practices that contribute to mistaken identification and how to reduce mistaken identifications by altering those practices (see Wells, Small, Penrod, Malpass, Fulero, & Brimacombe, 1998). The current work carries a somewhat different flavor. Although still emphasizing how law enforcement practices might be contributing to an eyewitness identification problem, the problem here concerns how certain practices could contribute to the loss of accurate identifications rather than the enhancement of mistaken identifications.

We are not yet prepared to argue that the use of composites should be significantly curtailed in criminal investigations. On the one hand, we know that composites usually produce results that poorly resemble the actual face that they were meant to depict (e.g., Ellis, Davies, & Shepherd, 1978; Ellis, Shepherd, & Davies, 1975; Christie & Ellis, 1981; Gibling & Bennet, 1984; Green & Geiselman, 1989; Kovera, Penrod, Pappas, & Thill, 1997) and that composites can apparently damage the memory that the witness has of the original face (current research). On the other hand, we do not know how often composites somehow manage to help crime investigators eliminate potential suspects or narrow the search of possible suspects. Furthermore, in multiple witness crimes it might be possible to use one witness to build a composite and save the other witnesses for any later lineup identification attempts. Furthermore, we cannot be certain that the effects we observed here are applicable to sketch artists. Like computer and transparency-based face composite systems, the sketch artist process is somewhat featural based rather than more purely holistic. Unlike computer and transparency-based face composite systems, however, the sketch-artist process does not require the witness to examine isolated facial features and select among them. Hence, we reserve judgment on the sketch artist process until we more fully understand the processes leading to the effects that we observed here.

References

Christie, D. F. M., & Ellis, H. D. (1981). Photofit constructions versus verbal descriptions of faces. *Journal of Applied Psychology, 66*, 358-363.

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 739-755.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior, 11*, 233-258.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition, 25,* 129-139.

Ellis, H. D., Davies, G. M., & Shepherd, J. W. (1977). Experimental studies of face identification. *Journal of Criminal Defense, 3*, 219-234.

Ellis, H. D., Shepherd, J. W., & Davies, G.M. (1975). An investigation of the use of the Photofit technique for recalling faces. *British Journal of Psychology, 69*, 467-474.

Fallshore, M., & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*, 1608-1623.

Gibling, F., & Bennett, P. (1994). Artistic enhancement in the production of Photofit likenesses: An examination of its effectiveness in leading to suspect identification. *Psychology, Crime, & Law, 1*, 93-100.

Green, D. L., & Geiselman, R. E. (1989). Building composite facial images: Effects of feature saliency and delay of construction. *Journal of Applied Psychology, 74,* 714-721.

Hosch, H. M., & Platz, S. J. (1984). Self-monitoring and eyewitness accuracy. *Personality and Social Psychology Bulletin, 10*, 283-289.

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.

Krafka, C., & Penrod, S. D. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology, 49*, 58-69.

Lindsay, R. C. L., Wells, G. L., & Rumpel, C. (1981). Can people detect eyewitness identification accuracy within and between situations? *Journal of Applied Psychology, 66*, 79-89.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35*, 409-420.

Loftus, E. F., Schooler, J. W., & Wagenaar, W.A. (1985). The fate of memory: Comment on McCloskey and Zaragoza. *Journal of Experimental Psychology: General, 114*, 375-380.

MacDonald, S. (1984, August 2). Wanted art: Meet a man who draws shady characters. *The Wall Street Journal*, pp. 1, 8.

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

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory-impairment hypotheses. *Journal of Experimental Psychology: General, 114*, 1-16.

Meisner, C., & Brigham, J. C. (2001). Twenty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*, 3-35.

Ryan, R. S., & Schooler, J. W. (1998). Whom do words hurt? Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology, 12*, S105-S125.

Sporer, S. L. (1996). Describing others: Psychological issues. In S. L. Sporer, R. S. Malpass, & G. Koehnken, *Psychological issues in person identification.*

Steblay, N. M., Dysart, J., Fulero, S., & Lindsay, R.C.L. (2003). Eyewitness accuracy rates in police showup and lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 27, 523-540.*

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 & Cognition, 25*, 583-592.

Wells, G.L. (1993). What do we know about eyewitness identification? *American Psychologist, 48*, 553-571.

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

Wells, G.L., Malpass, R.S., Lindsay, R.C.L., Fisher, R.P., Turtle, J.W., & Fulero, S. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist, 55*, 581-598.

Wells, G. L. & Olson, E. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied, 8*, 155-167.

Wells, G.L., Rydell, S.M., & Seelau, E.P. (1993). On the selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78*, 835-844.

Wells, G.L., Small, M., Penrod, S., Malpass, R.S., Fulero, S.M., & Brimacombe, C.A.E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads, *Law and Human Behavior, 22*, 603-647.

Wells, G.L., & Turtle, J.W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99*, 320-329.

Wells, G.L., & Windschitl, P.D. (1999). Stimulus sampling in social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115-1125.

Author Note

Gary L. Wells, Psychology Department; Steve D. Charman, Psychology Department; Elizabeth A. Olson, Psychology Department.

This research was supported by a grant from the National Science Foundation to the first author.

Correspondence concerning this article should be addressed to Gary L. Wells, Psychology Department, W 112 Lago Hall, Iowa State University, Ames, Iowa, 50011. Electronic mail may be sent to glwells@iastate.edu .

Footnotes

[1]There are too few controlled studies of sketch artists at this point to know whether they fare appreciably better than composite production systems in yielding likenesses of the intended target face. Such studies are made difficult by numerous factors, including the need to find and test representative samples of sketch artists (rather than one of two sketch artists). Accordingly, our discussion and conclusions in this article are restricted to facial composite systems and are not necessarily meant to generalize to sketch artists.

Table 1

Percentages of Participants Identifying the Target, Identifying a Filler, Making No Identification, and Identifying the Target When Forced to Make an Identification by Condition in Experiment 1

Condition	Identification of target	Identification of filler	No identification	Identification of target under forced identification
Control	84_a	6_a	10_a	94_a
Composite build	10_b	30_b	58_b	30_b
Yoked composite-exposure	44_c	6_a	50_b	82_a

Note: Percentages within columns not sharing a common subscript differ at $p < .05$ using a Chi Square test.

Table 2

Percentages of Participants Identifying the Target, Identifying a Filler, Making No Identification, and Identifying the Target When Forced to Make an Identification by Condition in Experiment 2

Condition	Identification of target	Identification of filler	No identification	Identification of target under forced identification
Control	60_a	4_a	36_a	88_a
Composite build	18_b	20_b	62_b	32_b

Note: Percentages within columns not sharing a common subscript differ at $p < .05$ using a Chi Square test.

Table 3

Percentages of Participants Identifying Someone or Making a Correct Rejection from Target-Absent Lineup in Experiment 2

Condition	Mistaken identifications	Correct rejections
Control	20	80
Composite build	26	74

Note: Percentages within columns do not differ significantly