Facial Composite Production by Eyewitnesses

Gary L. Wells and Lisa E. Hasel

Iowa State University

ABSTRACT—*The creation of facial images by eyewitnesses using composite-production systems can be important for the investigation of crimes when the identity of the perpetrator is at issue. Despite technological advances, research indicates that composite-production systems produce poor likenesses of intended faces, even familiar faces. Furthermore, producing a composite appears to harm later recognition performance. Although morphing composites from multiple witnesses helps, likeness is still limited. The problem might stem from a mismatch between how faces are represented in memory (holistically) and how composite systems attempt to retrieve the memories (at the feature level). New methods of face recall involving judgments of whole faces hold greater promise.*

KEYWORDS—*eyewitnesses; composites; face recall*

Kirk Bloodsworth, a U.S. military veteran living in Maryland, had never been in trouble with the law. Nevertheless, he was convicted of the 1984 rape and murder of a 9-year-old girl and was sentenced to die in Maryland's gas chamber. Bloodsworth became a suspect in the murder because an anonymous person called police to say Bloodsworth looked like a composite face that police had released to the media. Lacking a compelling alibi for the time of the crime, police placed his photo in a photo lineup and eyewitnesses identified him. After 9 years in prison, DNA evidence vindicated Bloodsworth and also implicated the actual murderer, Kimberly Ruffner (Junkin, 2004). Ruffner did not look much like the composite, but Bloodsworth did, and Bloodsworth was the only one in the photo lineup with hair that matched the composite.

The Bloodsworth case illustrates some key points in this article. First, laboratory research shows that a face composite by an eyewitness is generally a poor representation of the original face. Hence, a composite has the potential to lead crime investigators away from the real perpetrator and toward an innocent person. Also, a composite can bias the eyewitness away from identifying the original face and toward a face that resembles the composite. We review research on face composites, explore the question of why people are not better at building them, examine a new approach to face recall, and underscore the need for psychological science to help address this important problem in the justice system.

Address correspondence to Gary L. Wells, Psychology Department, West 112 Lago, Iowa State University, Ames, IA 50021; e-mail: glwells@iastate.edu.

COMPOSITE-PRODUCTION SYSTEMS

The objective of many criminal investigations is to establish the identity of the perpetrator. When there is a suspect, eyewitnesses can help establish identity by viewing a lineup that contains that suspect. When there is no suspect in the case, however, investigators often rely on eyewitnesses to help produce a likeness of the perpetrator's face. The first method for having eyewitnesses produce a face from memory was the sketch artist. Today, however, U.S. law enforcement agencies use mechanized systems and are over twice as likely to use computerized versions than noncomputerized ones (McQuiston-Surrett, Topp, & Malpass, in press).

Studies comparing sketch artists to mechanized systems are rare, perhaps because sketch artists vary widely in their skills and, hence, it would take a large sample of sketch artists randomly sampled from an ill-defined population to make conclusions. Accordingly, we restrict our review and discussion to composite-production systems such as the Identi-Kit, Photo-Fit, E-Fit, Mac-a-Mug, and FACES. The first two are early, noncomputerized collections of facial features (e.g., noses, eyes, mouths, head shapes, hair styles) that can be superimposed to create a face. The latter three are examples of modern, computerized versions of the same idea but include more possible facial features and more realistic visual results. FACES, for example, includes 361 hair selections, 63 head shapes, 42 forehead lines, 410 eyebrows, 514 eyes, 593 noses, 561 lips, 416 jaw shapes, 145 moustaches, 152 beards, 33 goatees, 127 eyeglasses, 70 eye lines, 147 smile lines, 50 mouth lines, and 40 chin lines. Figure 1

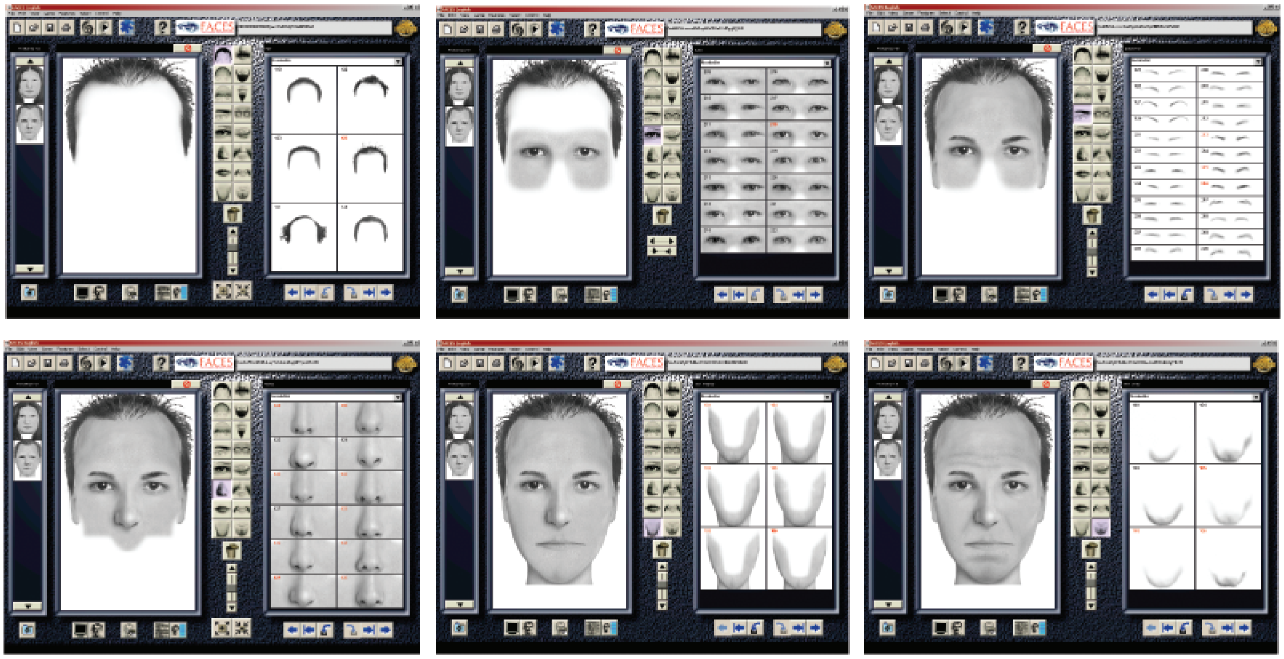

Fig. 1. Progressive sequence of building a composite face using the FACES 3.0 (IQ Biometrics) program. The eyewitness begins with any feature and adds features in any order. In this case, hair (upper left) is followed by eyes (upper middle); head shape and eyebrows (upper right); nose (lower left); lips and jaw shape (lower middle); and forehead lines, eye lines, smile lines, mouth lines, and chin lines (lower right).

shows a potential composite-building sequence using the FACES program. Each panel shows one or more features added in a sequence of steps. The eyewitness can begin with any feature and add features in any order. The upper left panel is hair only; the upper middle panel adds eyes; the upper right panel adds head shape and eyebrows (two steps); the lower left panel adds a nose; the lower middle panel adds lips and jaw shape (two steps); and the lower right panel adds forehead lines, eye lines, smile lines, mouth lines, and chin lines (five steps). A witness can go back and change any feature (e.g., eyes) by selecting a different one and the new feature replaces the old one.

The evolution of face-composite systems has generally involved increasing the number of available features, increasing the realism of the composite produced, and increasing the user-friendliness of the system. Recently, a new generation of face-recall systems has emerged that moves away from the feature-selection method of previous composite systems and instead uses whole faces. In one sense, these new systems, discussed later in this article, are not really composite systems at all, but they might hold greater promise for face recall than that of their predecessors.

HOW GOOD ARE COMPOSITES?

Research on the ability of people to use composite systems to produce likenesses of intended faces has a 30-year history, beginning with researchers in the United Kingdom. The results were disappointing from the outset (e.g., see Davies, Ellis, & Shepherd, 1978), and modern, computerized versions have not fared much better (Davies & Valentine, 2006). Many methods of assessing the accuracy of face composites have been developed, including matching tasks ("Which of these faces was this composite intended to depict?"), naming tasks (e.g., "What famous person is this?"), and similarity-rating tasks ("How similar is this composite to this actual face?"). One of the problems is the difficulty of comparing these measures across different studies, because each measure tends to depend on context. Matching tasks, for example, are sensitive to the similarity among the alternatives presented and the set size of the alternatives. Even if chance were defined as 25% in a four-alternative measure, a given composite might show a 60% match rate for four alternative faces if the faces were very different from one another or a 30% match rate if the alternative faces were more similar to each other.

If it is difficult to compare measures across studies, it is even more difficult to describe some "absolute" level of performance. So, why have researchers tended to conclude that composite-production systems typically produce poor likenesses of the intended targets? Poor relative to what? Clearly, this is a judgment call, but we agree with this overall conclusion based on two general observations. First, studies in which individuals attempted to create composites of famous faces show that very few people recognize the composite face even though they know the face of the famous person very well. Frowd et al. (2005), for instance, found that participants correctly named only 2.8% (22 of 800 possible) of a group of 50 famous faces based on

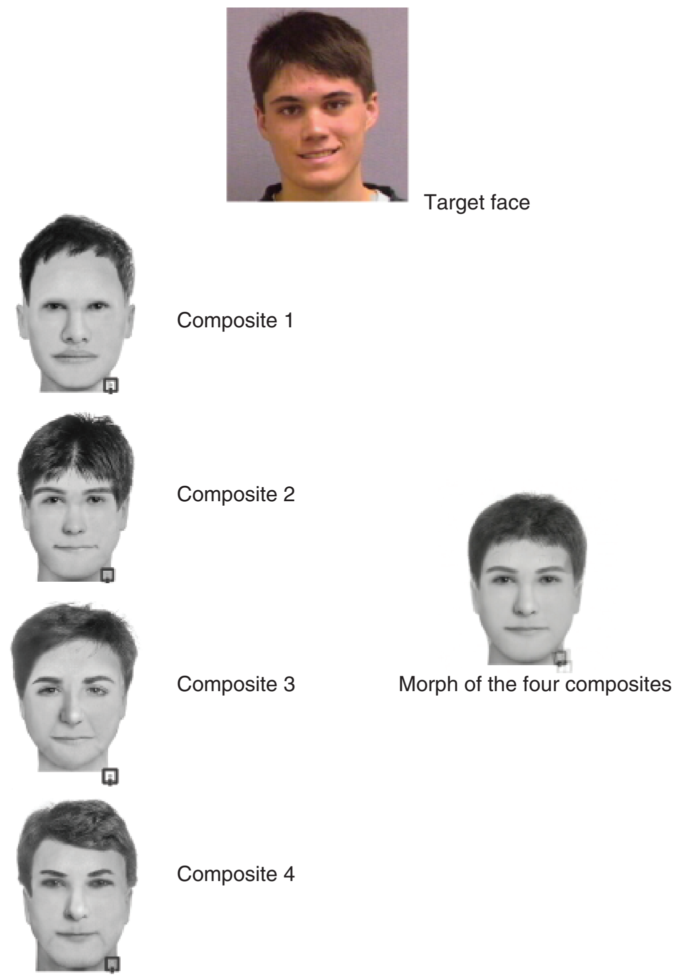

Fig. 2. A target face, four composites made by separate individuals, and the morph of the four composites. From the Hasel & Wells (2007) study.

composites of those faces that had been created by other individuals. Similarly, people seem largely unable to recognize people they personally know from composites of those individuals. Kovera, Penrod, Pappas, and Thill, (1997), for example, found that people were unable to discriminate between composites of their classmates they knew and composites of students from other schools whom they did not know. In general, studies of composite production reveal that participants create poor likenesses of intended faces, regardless of whether the composite builder knows the target face or has had only a brief encounter with that person, and regardless of the type of task (naming, matching, or similarity rating) used to assess the likeness (see Davies & Valentine, 2006, for an excellent review).

Recent research has shown that there can be benefits to morphing (averaging at the pixel level) multiple composite faces of the same individual created by different eyewitnesses (Bruce, Ness, Hancock, Newman, & Rarity, 2002; Hasel & Wells, 2007). Sampling across 16 different strategically sampled target faces, Hasel and Wells (2007) found that morphing composites from four independent participants, each of whom had viewed the same target face, produced a new image that was a better likeness of the target than was the average individual composite. Judges rated the similarity of each composite to the target face plus three distractor faces. Using the individual composites, the target received the highest rating 35% of the time (chance = 25%). Using morphed composites, however, the target received the highest rating 48% of the time. The superiority of the morph is evident in Figure 2, which depicts an example from the Hasel and Wells experiment. The figure shows one of the 16 target faces, four composites prepared by separate participants, meant to depict that target, and the morphed result from the four composites. The example in Figure 2 is one of the better results across the 16 target faces, but it illustrates the general pattern we observed: Morphing the composites produces a better likeness of the target than any one of the individual composites. Participants who built composites also rated how good they thought their own individual composites were. Those receiving the highest ratings were not more similar to the target faces than were the average individual composites, indicating that the composite builders were poor judges of how good or bad their own composites were.

WHY ARE COMPOSITES POOR REPRESENTATIONS OF THE INTENDED FACE?

No composite system could have enough facial-feature choices to represent the physiognomic variability of the human face. But, is the absence of enough features the problem with composite systems? Not likely. Wells and Hryciw (1984), for example, showed participants target faces that were created using the Identi-Kit and the participants also used the Identi-Kit for building their composites. Hence, it would have been possible to perfectly reconstruct the target face because every one of its features was available for selection from the composite kit. Nevertheless, resemblance scores from judges averaged less than 2.0 on a 7-point scale (1 = does not resemble, 7 = closely resembles). Furthermore, despite incredible advances in the technology of composite systems, including large increases in the number of facial features available, enhanced realism such that composites are almost indistinguishable from actual faces, and varied methods of approaching the task of building the composite, these technological advances have resulted in no consistent improvement in composite likenesses to the original faces (Davies & Valentine, 2006).

The dominant view in the research literature for over two decades has been that there is a mismatch between the task demands that are somewhat inherent in all composite systems and the way that faces are usually perceived and remembered. Numerous lines of evidence converge on the view that faces are generally processed, stored, and retrieved at a holistic level rather than at the level of individual facial features (see Tanaka & Farah, 2003). There are various views of what is meant by holistic, but the general idea is that the psychological process is not merely a processing of individual facial features (e.g., eyes, nose, mouth). Instead, faces might be represented in terms of their multidimensional similarity to other known faces (e.g., Valentine, 1995) or in a coordinate spatial relations system that includes distances between features, relative sizes of features, and so on that cannot be separated from the features themselves (Cooper & Wojan, 2000). As a result of holistic processing, recall of individual facial features is rather poor; and yet face-composite systems require the individual to recall individual facial features.

This is not to say that whole faces cannot be processed at the feature level if the task demands it. Wells and Hryciw (1984) had participants evaluate faces for 10 personality traits (e.g., how honest is this person?), a task that is likely to be performed based on whole-face processing, or evaluate faces for 10 physical-feature characteristics (e.g., how large is this person's nose?), a task that forces processing of isolated facial features. Later, participants either attempted to recognize the faces they had evaluated from six-person photo lineups or use the Identi-Kit to build the faces from memory. Those who rated the faces for personality traits performed much better on the lineup recognition tasks than did those who rated the individual features, but the opposite result occurred for the Identi-Kit task. This result is consistent with the idea that holistic processing (encouraged by personality-trait encoding) helps later recognition of the whole face but harms composite task performance whereas feature-based processing harms later recognition of the whole face and helps composite task performance.

A NEW SYSTEM APPROACH TO FACE RECALL

The view that faces are normally processed and stored in some type of holistic fashion rather than stored as constituent parts has led researchers to develop whole-face methods for face recall. These systems are still under development. They begin by generating a random set of faces and the witness selects the face most similar to memory for the target face. This face becomes the "parent" face that yields a new set of faces that are mutations of that face, using any of a number of possible algorithms, and the witness again makes a choice. This process is repeated until the witness cannot choose because all the faces resemble the target face equally well. Systems of this sort have been developed by Hancock (2000) and by Gibson, Pallares-Bejarano, & Solomon (2003). Comparison of these systems to traditional composite systems has thus far been very limited. The few tests thus far have not shown these particular versions of whole-face systems to be superior to traditional composite systems (Davies & Valentine, 2006). Nevertheless, these whole-face systems represent a radically different approach to producing composite faces and seem to hold the best prospects for a breakthrough because they rely on the idea, derived from basic research on face processing, that faces are processed holistically.

FINAL COMMENTS

The difficulty that people have with constructing face composites is not the result of weak memory for faces per se. People produce poor composite likenesses even for faces that they know very well and can easily recognize (Frowd, et. al., 2005). Instead, it appears that human face processing is designed more for face recognition, which is facilitated by holistic representations, than it is for face recall, which requires individual feature representations. Within a few days of birth, babies can recognize their mothers' faces, infants as young as 3 months show evidence of integrating facial features into a whole rather than perceiving them as individual features, and early visual experience appears to naturally set up a neural substrate for holistic processing of faces (Le Grand, Mondloch, Maurer, & Brent, 2004).

Why would human development favor a holistic representation of human faces rather than a feature-based representation of human faces? One possibility stems from our earlier observation that holistic representations facilitate recognition whereas feature-based representations facilitate recall (Wells & Hryciw, 1984). We speculate that evolutionary pressures might be responsible for mental systems favoring face recognition (and, hence, holistic processing) over feature-based face recall. Specifically, survival likely favored those who could readily recognize faces so as to make rapid judgments of a familiar versus unfamiliar face, friend versus foe, family versus nonfamily, and so on. But what was the survival value of face recall? It could be argued that face recall would have had some survival value (e.g., to communicate to family members the identity of a specific dangerous individual). However, a more efficient and reliable survival mechanism would be to simply avoid strangers, which favors face recognition and holistic processing over face recall and feature-based processing. Evolutionary processes that shaped the human brain could not anticipate modern technology and the needs and demands of a modern society.

Future research should give more weight to the question of how people naturally process faces so as to create face-recall systems that are congruent with natural face processing. Natural face processing has a strong holistic bias and, therefore, systems that require retrieval of memory for isolated facial features are not likely to ever work well. Creating systems that reflect how people actually process faces will require a better understanding of what is meant by holistic face processing and why holistic processing reduces the accessibility of information about individual features.

Attempts to evaluate and improve face composites fit into a larger problem in the criminal justice system. Analyses of the first 180 DNA exonerations in the United States reveal that mistaken eyewitness identification testimony was involved in 75% of the cases. Increasingly, it is becoming clear that errors in human memory are accounting for more of the convictions of innocent people than are all other causes combined. As the historical and natural home of the science of memory, psychological science has great promise for helping to solve an age-old problem.

Recommended Reading

Davies G.M., & Valentine T. (2006). (See References)

Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S. (2005). (See References)

Wells, G.L., Memon, A., & Penrod, S.D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45–75.

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

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

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.

Cooper, E.E., & Wojan, T.J. (2000). Differences in the coding of spatial relations in face identification and basic-level object recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 470–488.

Davies, G.M., Ellis, H.D., & Shepherd, J.W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, *63*, 180–187.

Davies, G.M., & Valentine, T. (2006). Facial composites: Forensic utility and psychological research. In R.C.L. Lindsay, D.F. Ross, J.D. Read, & M.P. Toglia (Eds.), *Handbook of eyewitness psychology Vol. 2* (pp. 59–86). Mahwah, NJ: Erlbaum.

Frowd, C.D., Carson, D., Ness, H., Richardson, J., Morrison, L., & McLanaghan, S. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, *11*, 33–52.

Gibson, S., Pallares-Bejarano, A., & Solomon, C. (2003). Synthesis of photographic quality facial composites using evolutionary algorithms. In R. Harvey & J.A. Bangham (Eds.), *Proceedings of the British Machine Vision Conference 2003* (pp. 221–230). London: British Machine Vision Association.

Hancock, P.J.B. (2000). Evolving faces from principal components. *Behavior Research Methods, Instruments and Computers*, *32*, 327–333.

Hasel, L.E., & Wells, G.L. (2007). Catching the bad guy: Morphing composite faces helps [electronic version]. *Law and Human Behavior*. Retrieved January 12, 2007 from www.springerlink.com.

Junkin, T. (2004). *Bloodsworth: The true story of the first death row inmate exonerated by DNA*. Chapel Hill, NC: Algonquin Books of Chapel Hill.

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.

Le Grand, R., Mondloch, C.J., Maurer, D., & Brent, H.P. (2004). Impairment in holistic face processing following early visual deprivation. *Psychological Science*, *15*, 762–768.

McQuiston-Surrett, D., Topp, L.D., & Malpass, R.S. (in presss). Use of facial composite systems in U.S. law enforcement agencies. *Psychology, Crime, and Law*.

Tanaka, J.W., & Farah, M.J. (2003). The holistic representation of faces. In M.A. Peterson & G. Rhodes. *Perception of faces, objects and scenes* (pp. 53–740). Oxford: Oxford University Press.

Valentine, T. (1995). *Cognitive and computational aspects of face recognition: Explorations in face space*. London: Routledge.

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

Author Query Form

Journal **CDIR**
Article **465**

Dear Author,

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers clearly on the query sheet if there is insufficient space on the page proofs. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Query No.	Description	Author Response
.	No Queries	