Check for updates

THE EFFECT OF FACIAL COMPOSITE CONSTRUCTION ON EYEWITNESS IDENTIFICATION ACCURACY IN AN ECOLOGICALLY VALID PARADIGM

GRAHAM E. PIKE
NICOLA A. BRACE
JIM TURNER
The Open University
ANNELIES VREDEVELDT
VU University Amsterdam

Previous research has produced equivocal results with regard to whether facial composite creation affects subsequent eyewitness identification accuracy, but the most widely publicized view is that creating a composite impairs the ability to later recognize the perpetrator from a line-up. In our first experiment, we examined this effect using several ecologically valid elements including a live staged crime, trained police officers, and a long delay between construction and identification, albeit with only a short delay between crime and composite construction. Composite construction did not significantly affect line-up identification accuracy. Experiment 2 replicated this result using a laboratory-based design and sequential line-up task, eliminating the possibly confounding effect of differential levels of motivation and relative judgments. Taken together, the experiments suggest composite creation may not negatively impact subsequent line-up accuracy, regardless of whether an ecologically valid method or more standard laboratory testing was used.

Keywords: criminal justice system; decision-making; memory; police; psychology

Despite advances made in forensic science, physical evidence features in only a small minority of criminal investigations. Indeed, Peterson, Sommers, Baskin, and Johnson (2010) estimated that forensic evidence linking a suspect to a crime scene is available in only about 2% of all cases. Instead, it is evidence obtained from eyewitnesses that plays a crucial role in criminal investigations. If the police do not have a suspect, eyewitnesses may be asked to construct a facial composite of the perpetrator. But once a suspect has been

AUTHORS' NOTE: *The authors would especially like to thank West Yorkshire Police, the Thames Valley Police Force, the Isle of Man Police Force, and the National Training Centre for Scientific Support to Crime Investigation, whose cooperation made this research possible. The research reported here was supported by the U.K. Home Office through the Innovations Fund Scheme.* *Correspondence concerning this article should be addressed to Graham E. Pike, School of Psychology, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK; e-mail: graham.pike@open.ac.uk.*

CRIMINAL JUSTICE AND BEHAVIOR, 2019, Vol. 46, No. 2, February 2019, 319–336.
DOI: 10.1177/0093854818811376
Article reuse guidelines: sagepub.com/journals-permissions
© 2018 International Association for Correctional and Forensic Psychology

found, those same eyewitnesses will often be asked to participate in a line-up identification procedure. It is, therefore, necessary to determine whether the construction of a facial composite affects witnesses' ability to later identify the perpetrator from a line-up.

As is discussed below, the results of research which has been conducted to answer this question have been somewhat equivocal. However, perhaps the most dramatic results were reported by Wells, Charman, and Olson (2005), who showed that composite production substantially impaired performance on a subsequent identification task. Their findings attracted considerable attention in the popular media (e.g., Moushey & Crabbe, 2005; Munger, 2006; Roth, 2007), in recommendations to police interviewers (e.g., Attorney General of the State of Wisconsin, 2005; Bockstaele, 2009), and in law review articles (e.g., McNamara, 2009; Trenary, 2013). This is clearly an important issue for psychologists, law enforcement personnel, and organizations that oversee or set policy and guidelines. The present article sought to contribute further to this important issue by investigating whether production of a facial likeness, using composite systems routinely used by the police, affects witnesses' ability to identify the perpetrator from a line-up several weeks later. Other ecologically valid elements included employing a live staged crime, composites constructed by trained police officers (with each officer only constructing a single composite), and a long delay between construction and identification. However, for practical reasons, it was only possible to use a shorter postcoding delay (i.e., between viewing the crime and constructing a composite) than would be present in an actual investigation.

The potential importance of witnesses constructing composites and attempting an identification at a subsequent line-up is perhaps best demonstrated through real cases. According to the latest inventory of DNA exoneration cases by the Innocence Project (2017a), 27% of eyewitness misidentifications involved facial composite sketches. Perhaps the most striking example is that of the first death row inmate exonerated by DNA evidence: Kirk Bloodsworth (Innocence Project, 2017b; Junkin, 2004). He became a suspect in a child rape and murder case because of his similarity to the composite created by eyewitnesses. Bloodsworth was put in a line-up, and three eyewitnesses identified him as the person they had seen with the victim (two additional eyewitnesses identified him later in court). Although this may sound like compelling evidence, the fact that these witnesses identified Bloodsworth from the line-up just proves that his appearance matched their memory of the perpetrator's appearance—which we already knew from the fact that Bloodsworth was apprehended based on that similarity. In real-world cases like this (and in any archival analysis), it is impossible to determine whether the witnesses' memory of the perpetrator's face was influenced by the process of constructing the facial composite. Instead, experimental research, that can establish cause and effect of specific factors, is needed to examine this potential influence.

The experimental research conducted to date has involved different facial composite procedures used by police, from precomputer systems, such as Identi-kit and Photo-FIT that use printed facial features, to computerized systems like E-FIT which use image files of facial features, to contemporary systems such as EFIT-V which use computer-generated faces. Jenkins and Davies (1985) investigated interference from viewing a presented composite on subsequent facial descriptions and identifications. After seeing a perpetrator, participants were shown a facial composite created with Photo-FIT software (a system that constructs the face by combining individual features from other faces that are selected by the witness), which was either "accurate" or altered (altered hair or added mustache). When later providing a description of the perpetrator, participants who had viewed the altered-feature

composite were significantly less accurate when describing that feature. A second experiment showed that participants who viewed the misleading composites were also significantly less accurate on a line-up task than those who had viewed no composite or an accurate composite. In a follow-up study, Gibling and Davies (1988) reported comparable impairments in recall and recognition as a result of viewing misleading composites. This impairment effect may be tempered, however, by allowing participants to view the target face for longer, or by asking them to recall the facial features of the target before viewing the misleading composite (Davies & Jenkins, 1985). Sporer (1996) showed participants either a "good" composite, a "misleading" composite, or no composite before performing an identification task and similarly found that participants who had viewed the misleading composite made significantly more "mix-ups." In contrast, Dekle (2006) reported that neither viewing a biased composite (which resembled a filler in the line-up) nor viewing an unbiased composite affected subsequent line-up identification accuracy.

Participants in the above-discussed studies were passive viewers of the composites, and the same effect may or may not occur for witnesses who are actively involved in the construction of a composite. The effect of active composite construction has been investigated in a range of studies that have also yielded mixed results. Some studies have suggested that constructing a composite benefits memory (Davis, Gibson, & Solomon, 2014; Mauldin & Laughery, 1981; McClure & Shaw, 2002; Meissner & Brigham, 2001), others have found that it harms memory (Comish, 1987; Kempen & Tredoux, 2012; Topp-Manriquez, McQuiston, & Malpass, 2016; Wells et al., 2005), and yet others have reported no significant effects (Davies, Ellis, & Shepherd, 1978; Davis, Thorniley, Gibson, & Solomon, 2016; Yu & Geiselman, 1993). As was mentioned earlier, the study conducted by Wells et al. (2005) received considerable publicity and attention, so in light of the impact of their results, we discuss Wells et al.'s study in more detail below.

In Wells et al.'s (2005) first experiment, participants viewed one of 50 photographed targets, provided a written description and then either constructed a likeness of the target using the FACES composite system, viewed a composite constructed by another participant, or performed no intervening task. Two days later, participants returned to perform an identification from a six-person photo line-up. The target face was present in all line-ups. Participants who had constructed a composite made far fewer correct identifications (10% vs. 84%) and far more foil identifications (30% vs. 6%) than participants who had performed no intervening task. In addition, those who had only viewed (not constructed) a composite made significantly fewer correct identifications (44%) than controls when given a free choice of response (i.e., allowed to say "not present"), although this difference disappeared when they were forced to make a selection. The authors concluded that passively viewing a composite does not disrupt memory; the lower free-choice correct identification rates in the passive viewing condition were interpreted as being due to participants' reduced confidence in their memory for the target, rather than their memory actually being impaired, because their forced-choice responses were as accurate as those of the control group.

Wells et al.'s (2005) second experiment replicated the first, except for the following differences: the passive viewing condition was omitted, exposure was to a single target face seen in a simulated-crime video, and both target-present and target-absent line-ups were included. For the target-present line-up, the results mirrored those of their first experiment: Participants who constructed a composite had considerably fewer correct identifications (18% vs. 60%) and more foil identifications (20% vs. 4%) than control participants. There

was no difference between groups for the target-absent line-up. The authors suggest that the latter result is due to the lack of any systematic bias in the line-up, as there would be no reason for an altered memory for the target (caused by constructing a composite) to resemble one line-up member more than any other.

Wells et al. (2005) explained the impairment they observed by pointing to the mismatch between featural and holistic processing. Specifically, the FACES composite system requires the witness to build a face by putting together individual features. This featural construction process conflicts with the holistic way in which humans naturally process and remember faces (e.g., Baddeley, 1979; Tanaka & Farah, 1993; Wilford & Wells, 2010). This was supported by the results of Davis et al. (2014; Davis et al., 2016), where a holistic system (EFIT-V) actually led to improved performance in some experiments.

However, the potentially damaging nature of featural construction has not been reported in all studies using featural composite systems, and some studies of Identi-kit (Mauldin & Laughery, 1981), E-FIT (Davis et al., 2014), and free-hand drawing (McClure & Shaw, 2002) have reported improvements in subsequent face recognition performance as well. This improvement could be due to a variety of mechanisms. If the composite is a good representation of the perpetrator's face, then it might help to preserve the memory of that face for retrieval at a later date. This is because composite construction essentially acts like rehearsal of accurate information, just like seeing a mugshot of the perpetrator would serve as a reminder of his appearance (cf. Lindsay, Nosworthy, Martin, & Martynuck, 1994). In that case, it would be best if witnesses construct the composite as soon as possible after the witnessed event. Alternatively, the process of composite construction might help witnesses to retrieve the image of the face from memory, just like mentally reinstating the context of the crime before viewing a line-up can help identification accuracy (e.g., Cutler, Penrod, & Martens, 1987; Krafka & Penrod, 1985; Malpass & Devine, 1981). In that case, it would be best if witnesses construct the composite just prior to attempting identification from a line-up. Mauldin and Laughery (1981) found support for the latter explanation. In their study, participants viewed a target and attempted identification 2 days later. The benefits of composite construction were greatest when construction occurred just prior to the identification task, rather than 2 days before identification. However, because the delay in real cases is commonly much longer than 2 days, any beneficial effect of composite production might be short-lived and thus confined to the laboratory.

The experiments conducted to date on facial composite construction have involved conditions that are far removed from conditions in most real-world criminal investigations. For example, none involved live events; they all involved viewing either pictures of faces (e.g., Davies et al., 1978; McClure & Shaw, 2002; Wells et al., 2005, Experiment 1) or a videotaped event (e.g., Davis et al., 2014; Davis et al., 2016; Wells et al., 2005, Experiment 2; Yu & Geiselman, 1993). Importantly, Ihlebæk, Løve, Eilertsen, and Magnussen (2003) reported that witnesses who viewed an event on video provided more complete and more accurate eyewitness statements than witnesses who experienced the same event live, suggesting that laboratory experiments may overestimate eyewitness memory performance in the real world. Furthermore, previous studies have typically involved unrealistically short delays between the witnessed event and creating the composite, and between creating the composite and viewing the line-up. Eyewitnesses in real life will typically not create a composite immediately after witnessing a crime. In a similar vein, after the composite has been created, it will take some time before the police are able to find and apprehend a suspect.

Finally, in previous experiments, the composite systems were typically operated by the participants themselves or by inexperienced operators (but see Davies et al., 1978; Davis et al., 2014, Experiment 2, for exceptions). In real life, witnesses will normally describe the perpetrator to a trained, experienced police operator, who will create the composite based on the witness' description. The goal of Experiment 1 was to approximate real-life conditions more closely.

EXPERIMENT 1

Although the differences in the results reported to date are intriguing and worthy of further exploration, perhaps the most pressing need is to establish whether facial composites affect subsequent identification accuracy in actual police investigations. One reason for Wells et al.'s (2005) second experiment was increasing the ecological validity of their materials, but this experiment still differed in some significant ways from what would happen in a police investigation, such as the use of a video-recorded event rather than a live event, the relatively short delay between creating the composite and viewing the line-up (48 hr), and the fact that participants operated the composite software themselves. Our first experiment was, therefore, designed to extend the scope of existing research by employing a more ecologically valid methodology. To this end, the target perpetrator was seen in a live staged event, simulating a minor (nonviolent) criminal offense; a realistic delay (4-6 weeks) was introduced between composite construction and identification; composites were created by trained policing personnel exactly as they would be in a criminal investigation; and the composite software used was a system commonly used by police in the United Kingdom (E-FIT). To enable comparisons with previous research, particularly that of Wells et al. (2005), we used a simultaneous photo line-up task, which would not be considered ecologically valid in comparison to identification procedures used in the United Kingdom, but does match procedures used outside the United Kingdom (notably the United States).

METHOD

Participants and Design

Eighty volunteers were recruited from the local communities of the National Training Centre for Scientific Support to Crime Investigation (County Durham) and the Douglas police station (Isle of Man). Three participants in the control condition who failed to return the completed line-up task within 6 weeks were excluded from all analyses. This left 77 participants (55% female), with a mean age of 38. To match the design of the second experiment reported by Wells et al. (2005), participants were randomly assigned to one of four conditions of a 2 (condition: composite building, control) $\times$ 2 (line-up: target-present, target-absent) factorial design. This sample size allowed us to detect a large effect (of $\phi = .50$ for the target-present condition and $\phi = .46$ for the target-absent condition) with power = .80 at $\alpha = .05$ (i.e., sufficient to detect the key effect of $\phi = .74$ found by Wells et al.).

Materials

The staged crime closely followed a predetermined script. The perpetrator, who was a Caucasian male, approximately 28 to 30 years old with short hair, no facial hair or glasses, and no other distinguishing characteristics, was unfamiliar to the witnesses and to the

policing personnel who worked with them to create a composite. He entered a car park and walked around two cars, before removing a video camera from the front passenger seat of a third. The cars were approximately 10 to 15 ft from the witnesses. At the start and end of the scene, which lasted about 60 s, the perpetrator paused to look around the car park, during which time the witnesses got a clear look at his face from a number of angles, for approximately 20 s and from less than 10 ft away.

Composites were created using E-FIT software, operated by a trained police officer. E-FIT consists of a large database of grayscale images of facial features and works by an operator inputting the description provided by the witness (obtained using a Cognitive Interview) into the system by selecting descriptors (from the Aberdeen Index of facial descriptors). This generates an initial likeness, after which operator and witness work together to change, move, and resize features (and can edit the image using standard image manipulation software) until the witness is satisfied that the image looks like the face of the perpetrator (for more information, see Fodarella, Kuivaniemi-Smith, Gawrylowicz, & Frowd, 2015). The operators in this experiment produced the E-FIT as their final assessment in a facial composite training course. Each participant worked with a different police operator (i.e., each operator interviewed one witness and constructed only one composite). The assessment process meant that an experienced operator verified that each operator had followed the prescribed interviewing and construction methods.

High-quality color photographs were used to create the line-ups. Target-absent line-ups consisted of nine foils who had a similar appearance to the perpetrator (i.e., same ethnicity, age, hair color, facial build, and no other distinguishing characteristics). For the target-present line-ups, one foil was randomly dropped and replaced with the target. The target, as in real investigations, had been seen in real life previously, so under different lighting and background conditions, and wearing different clothes compared with the image used in the line-up. All line-ups consisted of a 3×3 grid of faces, with a number label (from 1-9, consecutively) placed under each face. The position of the target was varied to produce nine versions (with the target in different locations), and the foil that was missing from the target-present line-ups was also varied across these nine versions, to minimize the difference between target-present and target-absent line-ups.

Procedure

To meet informed consent requirements, all participants were told that they would view a staged, nonviolent crime, and would later be asked to recall details of what happened. They were not informed that they would be asked to identify the perpetrator at a later date. The staged crime was repeated 4 times, always within a policing facility, with each occasion viewed by 18 to 28 witnesses. On arrival, witnesses were told not to confer with one another before, during, or after the staged crime, and a researcher stayed with the group of participants to ensure that this instruction was obeyed. After the staged crime, participants were divided into composite construction and control groups. The composite construction participants were led to individual offices and introduced to the police officer who would be working with them to construct the composite (this process took 10-15 min), while control participants were asked to wait for 10 to 15 min before providing their verbal description. The current study followed Wells et al.'s (2005) design, in that participants in both conditions were asked to individually provide a description. Participants were asked to recall all

TABLE 1: Target-Present and Target-Absent Line-Up Outcomes (%) by Condition in Experiment 1

	Target-present		Target-absent	
Line-up outcome	Composite (n = 20)	Control (n = 19)	Composite (n = 20)	Control (n = 18)
Hit	70.0% (14)	63.2% (12)	NA	NA
False alarm	10.0% (2)	10.5% (2)	25.0% (5)	38.9% (7)
Miss	20.0% (4)	26.3% (5)	NA	NA
Correct reject	NA	NA	75.0% (15)	61.1% (11)

Note. Absolute numbers are shown in parentheses.

they could about the staged crime and perpetrator. After the free-recall description, they were prompted to provide information about each facial feature. Following this, control participants were escorted from the policing facility, while composite construction participants created an E-FIT image with the police officer (see "Materials" section for details). Composite construction took approximately 60 min on average.

Four weeks after viewing the staged crime, participants received the photo line-up, instructions, and a return envelope in the post. The line-up instructions stated,

> The person you saw committing the crime MAY or MAY NOT be one of the men in the photographs. If you think that one of the pictures is that of the perpetrator, please write the number under the image in the box below. Please write only one number. If you do not think that any of the photographs are that of the perpetrator, please write "NO" in the box.

Participants indicated their confidence in this decision on a Likert-type scale ranging from 1 (*very unconfident*) to 6 (*very confident*). Seventy-seven completed line-ups were returned: 40 from the composite construction participants and 37 from the control participants. The delay between the event and receipt of the completed line-up ranged from 29 to 42 days (M = 36 days, SD = 3 days).

RESULTS

The data from Experiment 1 are presented in Table 1. For target-present line-ups, there were three possible outcomes: a "hit," correctly identifying the perpetrator; a "false alarm," where they identified one of the foils; and a "miss," where they did not select anyone from the line-up. The data in Table 1 show that most participants chose someone from the target-present line-up rather than not choosing anyone, both in the composite building condition (80% choosers and 20% nonchoosers) and the control condition (73.7% choosers and 26.3% nonchoosers). The data from the target-present line-ups were analyzed using a 3 (outcome) $\times$ 2 (condition) chi-square test, which showed a statistically nonsignificant result with a small effect size, $\chi^2(2) = 0.24$, $p = .887$, $\phi_c = .08$. Thus, participants in the composite condition were no more likely to make a correct or foil identification than participants in the control condition. For target-absent line-ups, there were two possible outcomes: a "correct rejection," where participants did not select anyone; and a "false alarm," where they identified one of the foils. Like the target-present data, the target-absent data for both the composite building and control conditions show a tendency for participants to be correct rather than incorrect, that is, to be nonchoosers rather than choosers. The data from the target-absent

TABLE 2: Mean Confidence Rating (1-6) by Condition and Line-Up Outcome in Experiment 1

	Target-present			Target-absent		
Condition	Hit	False alarm	Miss	Reject	False alarm	Total
Composite	4.9 (1.3)	1.5 (0.7)	3.0 (0.8)	5.2 (1.0)	4.4 (2.1)	4.6 (1.5)
Control	5.1 (0.8)	3.0 (1.4)	2.2 (1.1)	5.1 (0.7)	4.0 (1.8)	4.4 (1.5)
Total	5.0 (1.1)	2.3 (1.3)	2.6 (1.0)	5.2 (0.9)	4.2 (1.9)	4.5 (1.5)

Note. Standard deviations in parentheses.

line-ups were analyzed using a 2 (outcome) × 2 (condition) chi-square test, which showed a statistically nonsignificant result with a small effect size, $\chi^2(1) = 0.85, p = .358, \phi_c = .15$.

In both experiments by Wells et al. (2005), the authors used four "orthogonal chi-square tests" (with an adjusted alpha value of .0125) and made estimates of effect size using phi. This was done to compare individual cells within the larger 3 × 3 (Experiment 1) and 2 × 2 × 3 (Experiment 2) contingency tables to determine more precisely where key differences occurred. For example, the hit rates for the "yoked" and "composite building" conditions in their Experiment 1 were compared in this way. To facilitate comparison with Wells et al., the present data were analyzed using three likelihood ratio tests (essentially the same as "orthogonal chi-square tests"), employing the technique described by Howell (2002). Following Wells et al., the standard equation for the phi coefficient, $\phi = \sqrt{(\chi^2/N)}$, was employed, although it is worth noting that this is suitable for 2 × 2 designs only. Phi coefficients of .1, .3, and .5 are considered small, medium, and large effects, respectively. Only three tests were conducted, as the current design did not include the additional "forced choice" task used by Wells et al. The first two analyses, on target-present line-ups, revealed no significant differences between conditions for either the number of correct identifications, $\chi^2(1) = 0.07, p = .385, \phi = .05$, or the number of foil identifications, $\chi^2(1) = 0.01, p = .920, \phi = .05$. The third analysis, on target-absent line-ups, revealed no statistically significant difference between conditions for the number of foil identifications, $\chi^2(1) = 0.57, p = .452, \phi = .22$. As stated by Wells et al., further analysis of the data pertaining to misses or correct rejections is not possible as these data are not orthogonal to analysis of hits and false alarms.

The confidence data (see Table 2) were analyzed using a 2 (condition: composite building, control) × 2 (line-up: target-present, target-absent) × 2 (decision: correct, incorrect) analysis of variance (ANOVA), which revealed a nonsignificant main effect of condition, $F(1, 69) = 0.13, p = .719, \eta_p^2 = .00$, a significant main effect of line-up type, $F(1, 69) = 10.10, p = .002, \eta_p^2 = .13$, and a significant main effect of decision accuracy, $F(1, 69) = 34.94, p < .001, \eta_p^2 = .34$. Participants were more confident for target-absent line-ups than for target-present line-ups and more confident in correct than in incorrect decisions. The two-way and three-way interactions were nonsignificant, except for the interaction between line-up type and decision accuracy, $F(1, 69) = 7.32, p = .009, \eta_p^2 = .10$. Table 2 shows that the difference in confidence between correct and incorrect decisions was greater for target-present line-ups, $F(1, 69) = 38.57, p < .001, \eta_p^2 = .36$, than for target-absent line-ups, $F(1, 69) = 4.95, p = .029, \eta_p^2 = .07$. There was a moderate and statistically significant point-biserial confidence-accuracy correlation in the composite building condition, $r_{pb}(40) = .50, p = .001$, and a strong, statistically significant correlation in the control condition, $r_{pb}(37) = .62, p < .001$.

DISCUSSION

The results of Experiment 1 provided no evidence that building a facial composite of a perpetrator affected participants' ability to later identify that perpetrator from a target-present line-up. Likewise, there was no evidence that participants who constructed a facial composite were more likely than the control group to identify the wrong person from a target-absent line-up. Indeed, although the differences between conditions were statistically nonsignificant with small effect sizes, participants in the composite building condition were actually slightly more likely to identify the target and less likely to make a misidentification than participants in the control condition. There were also no significant differences in confidence between conditions, but participants who made a correct decision were significantly more confident than those who were incorrect, particularly for target-present line-ups. Finally, there was a significant correlation between confidence and accuracy, suggesting that participants were generally able to monitor the accuracy of their decisions.

This pattern of results is very different to that reported by Wells et al. (2005), but similar to the null findings reported in other studies (Davies et al., 1978; Davis et al., 2016; Yu & Geiselman, 1993). There are a number of differences between the current experiment and that of Wells et al. that could explain the variations in the results, which will be discussed in detail below. However, it is also worth noting that Wells et al.'s second experiment had more than twice as many participants as our Experiment 1, so one possibility is that the experiment reported here did not have sufficient statistical power to reveal a significant impairment. However, given that the trend in our data is in the opposite direction from that in Wells et al., a lack of statistical power could only be seen as preventing a facilitating effect from having been observed, rather than masking any interference effects. Other than number of participants, Experiment 1 also differed from the second, more ecologically valid, experiment conducted by Wells et al. (2005) in terms of using a live, rather than videotaped event; employing a delay of 4 to 6 weeks, rather than 2 days; using the E-FIT system rather than the FACES program; employing trained policing personnel to construct the composites rather than the participants constructing the composite by themselves; and using a nine person line-up, rather than a six-person line-up. Although it is possible that any of these factors, or indeed combinations of these factors, could explain the differences in results, one possible interpretation is that any deleterious effects on identification of creating a facial composite are not apparent when particularly realistic conditions are used. Obviously, this is of particular relevance to research that is addressing an applied question.

One likely effect of increasing ecological validity might be to also increase perceived consequentiality. In terms of actual consequentiality, both our Experiment 1 and the experiments of Wells et al. were essentially identical. However, asking participants to view a live event at a police facility and to be interviewed by policing personnel might well have led participants to engage more seriously or realistically with the task than would students in a laboratory. Such an increase in perceived consequentiality should not lead to a change in the pattern of results, merely their magnitude. However, it is also possible that participants in the composite building condition in Experiment 1 perceived greater consequentiality than not only the participants in Wells et al.'s second experiment but also participants in the control condition as well, given that they spent considerable time working with and being interviewed by a police officer, whereas participants in the control condition did not. This greater level of perceived consequentiality might have motivated them to engage with the task more and this greater motivation could explain why they performed (albeit nonsignificantly)

better than the controls and have masked a similarly deleterious effect to that observed by Wells et al.

EXPERIMENT 2

To minimize any difference in perceived consequentiality and therefore motivation between the composite building and control conditions, Experiment 2 replicated Experiment 1 in a laboratory setting. We used a staged criminal offense viewed on video, rather than a live event, and participants in the composite building condition worked in a laboratory setting with a researcher, rather than with policing personnel in a policing facility. In addition to reducing ecological validity, two other changes were made. The first was to reduce the delay between composite building/verbal description and the subsequent line-up task from 4 to 6 weeks to 7 to 10 days. This is practically more feasible but still falls within the range of delays likely to occur in a real police investigation.

The second change involved altering the format in which the line-up was conducted. Experiment 1 used a simultaneously presented photo array, to replicate the technique used by Wells et al. (2005). It is likely that such a line-up would have encouraged participants to use relative judgments (Wells et al., 1998). Although the same line-up task was used for both groups, there is a possibility of an interaction between condition and task, such that participants in the composite building condition were advantaged. For example, Wells and Hryciw (1984) found that participants who had focused on featural information while encoding a face (by rating facial features) performed better on a subsequent featural retrieval task (constructing a composite) than participants who had focused on holistic information while encoding (by rating personality traits). In a similar vein, one could argue that the great many relative judgments made during the composite construction process in Experiment 1 could have primed relative processing on the simultaneous line-up, and as such improved performance for the composite construction group, or at least masked any deleterious effects of construction. Although such priming effects are likely to be short term, and certainly not expected to prime processing 4 to 6 weeks later, we limited the use of relative judgments during retrieval by introducing a sequentially presented line-up in Experiment 2. In this case, a video-recorded "VIPER" line-up was employed, a procedure which is currently the preferred method of identification by police in the United Kingdom. A brief description of the VIPER line-ups used in this experiment is given in the "Method" section below (for a more detailed description of the VIPER system, see Kemp, Pike, & Brace, 2001).

METHOD

Participants and Design

Seventy-four participants volunteered to take part. One participant in the control condition who failed to complete the second half of the experiment was excluded from all analyses. This left 73 participants (81% female with a mean age of 37). All were working at the main campus of the university in a variety of capacities, including secretarial, administrative, and academic posts. They had not taken part in previous research involving composite construction. Participants were randomly assigned to one of four conditions of a 2 (condition: composite building, control) × 2 (line-up: target-present, target-absent) factorial

design. This sample size allowed us to detect a large effect (of $\phi = .51$ for the target-present and $\phi = .47$ for the target-absent conditions) with power = .80 at $\alpha = .05$.

Materials

A video sequence of a staged crime scenario was filmed, depicting two Caucasian males fighting over a bag and shouting at each other. The sequence was edited to ensure that one of the two males was viewed from close-up as well as from a distance and from all angles, and then edited again to produce a second version of the film in which the other male featured prominently. Thus, two films were created featuring two different targets, but with very similar events. Participants were randomly assigned to a film version. The edited sequences were just more than 1 min in length. A target-absent and a target-present video parade were prepared by the West Yorkshire police, following standard police guidelines for the construction of video parades and using their VIPER system. Each parade consisted of nine head-and-shoulder shots presented in a sequence of video clips, each lasting for 15 s with an interstimulus-interval (consisting of a black screen) of 3 s. Each face was shown facing forward for 5 s, then turning to the right profile for 5 s, and then to the left profile for 5 s. The faces were captured and presented in high-resolution color, filmed against a standardized background and using standardized lighting, and robustly matched for image properties.

Procedure

All participants were informed that they were about to act as witnesses and would view a short video of a staged "incident" involving two men. They were told that after seeing the video they would be asked to provide as complete a verbal description as possible of one of the men, feature by feature. Participants in the composite construction condition were also told that after their description they would work with the researcher (who had been trained in using the E-FIT system and had constructed several dozen composites before but was not a police E-FIT operator) to produce a facial composite of one of the men using the E-FIT system. Participants then individually watched the video and, after a short filler task unrelated to the experiment, were given further instructions to write down as much as they could about the target including clothing, body, and facial details. The free-recall phase was followed by prompts for each facial feature. Once they had completed their description, participants in the control condition ($N = 36$) left. Participants in the composite construction condition ($N = 37$) were informed that there would be a short break while the details recalled were entered into a facial composite system. Once an initial composite image was generated the participants were invited to work with the interviewer to alter the composite until they felt they had achieved a likeness to the target that was as good as possible.

Between 7 and 10 days later, participants returned to the laboratory. They were reminded that about a week ago they had acted as a witness to a staged incident shown on video and, depending on experimental condition, that they had provided either a verbal description only or a verbal description and a facial composite of one of the men depicted in the video. They were asked to attempt to identify that man from a video identification parade. It was explained that the parade would consist of nine head-and-shoulder shots with a number above it and that each parade member would be shown both from the front and from each side. They were informed that the person that they had seen previously may or may not be

TABLE 3: Target-Present and Target-Absent Line-Up Outcomes (%) by Condition in Experiment 2

	Target-present		Target-absent	
Line-up outcome	Composite (n = 19)	Control (n = 18)	Composite (n = 18)	Control (n = 18)
Hit	57.9% (11)	44.4% (8)	NA	NA
False alarm	15.8% (3)	22.2% (4)	16.7% (3)	11.1% (2)
Miss	26.3% (5)	33.3% (6)	NA	NA
Correct reject	NA	NA	83.3% (15)	88.9% (16)

Note. Absolute numbers are shown in parentheses.

present in the video parade, and that they would see the entire parade through twice before making a decision (in line with U.K. Police and Criminal Evidence Act codes of practice). They were instructed to indicate to the researcher either the number of the face they believed to be that of the perpetrator, or that they did not think that any of the images were of the person. They rated how confident they felt in their decision on a 6-point Likert-type scale. For both conditions, half of the parades shown were "target-present" (a clip of the target was included in the parade) and half "target-absent" (there was no clip of the target). Line-up administrators were blind to experimental condition and position of the target. Finally, participants were debriefed and asked not to discuss the study with others.

RESULTS

The data for target-present and target-absent line-ups in Experiment 2 are presented in Table 3, using the same outcomes as in Experiment 1 (hits, false alarms, misses, and correct rejections). As there were no significant differences between the two versions of the film for either line-up accuracy, $\chi^2(1) = 0.70, p = .404, \phi_c = .09$, or confidence, $t(71) = 0.05$, $p = .963$, $d = .01$, data from both versions were pooled for all subsequent analyses. The data in Table 3 show that most participants chose someone from the target-present line-up rather than not choosing anyone, both in the composite building condition (73.7% choosers and 26.3% nonchoosers) and the control condition (66.6% choosers and 33.3% nonchoosers). A 3 (outcome) × 2 (condition) chi-square test revealed a statistically nonsignificant result with a small effect size, $\chi^2(2) = 0.68$, $p = .711$, $\phi_c = .14$. Thus, participants in the composite condition were no more likely to make a correct or foil identification than participants in the control condition.

The target-absent data for both the composite building and control conditions displayed in Table 3 show a tendency for participants to be correct (86.1% overall) rather than incorrect (13.9% overall). A 2 (outcome) × 2 (condition) chi-square test showed a statistically nonsignificant result with a small effect size, $\chi^2(1) = 0.23$, $p = .630$, $\phi_c = .08$. As in Experiment 1, these data were also analyzed using three likelihood ratio tests to facilitate comparison with Wells et al.'s (2005) analyses using orthogonal chi-square tests. The two analyses conducted on target-present line-ups revealed no statistically significant differences between conditions for either the number of hits, $\chi^2(1) = 0.48$, $p = .490$, $\phi = .16$, or the number of false alarms, $\chi^2(1) = 0.14$, $p = .705$, $\phi = .14$. The third analysis, on target-absent line-ups, also revealed no statistically significant difference between conditions for the number of false positives, $\chi^2(1) = 0.02$, $p = .654$, $\phi = .20$.

TABLE 4: Mean Confidence Rating (1-6) by Condition and Line-Up Outcome in Experiment 2

	Target-present			Target-absent		
Condition	Hit	False alarm	Miss	Reject	False alarm	Total
Composite	4.7 (1.3)	2.1 (1.5)	4.4 (1.1)	4.7 (1.0)	4.0 (0.0)	4.4 (1.2)
Control	4.4 (0.7)	4.3 (0.5)	4.0 (1.1)	4.7 (0.7)	3.5 (0.7)	4.4 (0.8)
Total	4.6 (1.1)	3.6 (1.3)	4.2 (1.1)	4.7 (0.9)	3.8 (0.4)	4.4 (1.0)

Note. Standard deviations in parentheses.

The confidence data, displayed in Table 4, were analyzed using a 2 (condition) × 2 (line-up type) × 2 (decision accuracy) ANOVA, which revealed a nonsignificant main effect of condition, $F(1, 65) = 0.16$, $p = .687$, $\eta_p^2 = .00$, a nonsignificant main effect of line-up type, $F(1, 65) = 0.01$, $p = .935$, $\eta_p^2 = .00$, and a significant main effect of decision accuracy, $F(1, 65) = 6.80$, $p = .011$, $\eta_p^2 = .09$. Participants were more confident in their correct decisions than in their incorrect decisions. The two-way and three-way interactions were nonsignificant. There were moderate and statistically significant point-biserial confidence-accuracy correlations in both the composite building condition, $r_{pb}(37) = .33$, $p = .044$, and the control condition, $r_{pb}(36) = .35$, $p = .038$. These were slightly lower than in Experiment 1.

DISCUSSION

The results of Experiment 2 were similar to those of Experiment 1: not only were there no deleterious effects of composite building, but the pattern of results was actually in the opposite direction (although statistically nonsignificant). Thus, even under conditions that reduced the perceived consequentiality of the experiment, and therefore presumably participants' motivational levels, our data revealed no significant differences in identification accuracy between conditions. Composite production similarly did not affect the level of confidence expressed in line-up decisions, nor the confidence-accuracy correlation. Overall, participants were more confident in correct than incorrect decisions and the data showed a moderate relationship between confidence and accuracy.

GENERAL DISCUSSION

Although the results of previous research on the effects of facial composite creation on identification accuracy have been mixed, the findings from one particular study (Wells et al., 2005) have been publicized widely. Their finding that composite building substantially impaired identification accuracy has been promoted among police practitioners (e.g., Attorney General of the State of Wisconsin, 2005), legal professionals (e.g., McNamara, 2009), and the public (e.g., Roth, 2007). In the two experiments presented here, however, we did not replicate that finding. We found no deleterious effects of composite building on the ability to identify the perpetrator from a target-present line-up or on the ability to correctly reject a target-absent line-up. Participants who had created a composite using the E-FIT system shortly after the incident performed just as well as (if not slightly better than) participants who had not created a composite.

Our first experiment was designed to approximate real-life conditions as closely as possible, and more closely than the experiments conducted by Wells et al. (2005). We exposed participants to a live staged crime, introduced a substantial delay between composite creation and line-up identification, and had trained police operators create the facial composites using E-FIT, a system used frequently by police in the United Kingdom. The crucial finding from this experiment was that composite creation in this highly realistic setting did not impair identification accuracy.

Of course, by improving ecological validity in Experiment 1, we also introduced various differences between Experiment 1 and Wells et al.'s experiments, making it difficult to pinpoint the reason for the conflicting results. Any deleterious effects of composite building could have been offset by two specific advantages that participants in the composite creation group in Experiment 1 had over the control group. First, participants in the composite condition may have been more motivated to perform well on the identification task, simply because they had spent more time working with a police operator at the police facility. Even though it is unlikely that simply trying harder will improve performance on a line-up task, since accurate recognition is associated with automatic rather than deliberative processes (Dunning & Stern, 1994), participants' motivation level remains a possible confounding factor. The second potential advantage was that the composite creation group had spent a lot of time comparing facial features to create the composite, and may therefore have been primed to use the type of relative judgment that facilitates identification from a simultaneous line-up (cf. Wells & Hryciw, 1984). Even though the intervening delay of 4 to 6 weeks would have likely eliminated any such priming effects, this also remains a potential confound. The main aim of Experiment 2 was to study the impact of composite creation when those two advantages were eliminated. Even in a laboratory experiment with sequential line-ups, the results still showed that composite creation had no significant effect on identification accuracy. Thus, perceived consequentiality and priming of relative processing cannot explain the absence of deleterious effects of composite construction.

Two other factors may have inoculated participants in our experiments to potential deleterious effects. First, participants in all conditions provided a detailed verbal description of the target shortly after seeing him. Initial recall attempts can protect against the incorporation of subsequently encountered misinformation (e.g., Gabbert, Hope, Fisher, & Jamieson, 2012; Memon, Zaragoza, Clifford, & Kidd, 2010). Davies and Jenkins (1985) found that an immediate attempt to recall the target's facial features, provided that the features were recalled correctly, eliminated the harmful effect of viewing a misleading composite. Second, unlike witnesses in real life, participants in both experiments created a composite fairly quickly after the incident. Thus, their memory for the target's face was still relatively strong, and therefore perhaps less sensitive to the potentially harmful effects of composite construction. Neither of these factors can provide a satisfactory explanation for the absence of a composite impairment effect in our experiments, however, since both the early opportunity for recall and the short delay were also present in Wells et al.'s (2005) experiments, which did find impairment effects. Furthermore, Davies et al. (1978) concluded that the length of the delay between encoding and composite construction did not affect the impact of composite construction on face recognition accuracy. Nonetheless, the short delay between the incident and composite construction is a limitation to the ecological validity of our study (cf. Frowd et al., 2005), and we recommend that future researchers explore the effects of a longer delay.

Having ruled out these factors, at least three potential explanations for the conflict between our results and those of Wells et al. (2005) remain. First, it is possible that our participants were adopting a different response criterion than those of Wells et al. This would be a particular problem if the higher hit rate in the composite building condition in our Experiment 1 was simply a product of participants being more likely to make a selection from the line-up. However, analysis of the data suggests this is not a viable explanation, first because the proportion of choosers versus nonchoosers was very similar in both conditions for the target-present line-ups and also because there was no significant difference between conditions for the target-absent line-ups.

Second, we used a longer delay between composite creation and identification (4-6 weeks in Experiment 1 and 7-10 days in Experiment 2) than Wells et al. (2 days in both experiments). Mauldin and Laughery (1981), who concluded that composite creation improved subsequent identification performance (i.e., the opposite effect to Wells et al.), showed that the length of delay affected the results. Composite creation in their study was most beneficial when it occurred 2 days after encoding (i.e., right before the identification task). It was still helpful when it occurred immediately after encoding (i.e., 2 days before the identification task), but to a lesser degree. If the trend observed by Mauldin and Laughery were to continue, one would expect composite construction to be even less helpful, or no longer helpful at all, if it occurred 7 to 10 days or even 4 to 6 weeks before identification. Indeed, our results showed no significant benefits after those longer delays. The findings of Wells et al., however, do not fit with this pattern: they observed impairment when composite creation occurred 2 days before identification (where Mauldin and Laughery had still found benefits, albeit smaller than with an even shorter delay). Thus, the variability in delays cannot explain why Wells et al.'s study revealed a disruptive effect of composite creation on identification accuracy whereas ours did not, even though they used a shorter delay.

Third, we used a different composite system than Wells et al. (2005). They suggested that the piecemeal processing of facial features, necessitated by building a composite using the FACES software, was a likely contributing factor to less accurate performance on the subsequent line-up task. Like FACES, the E-FIT system also requires that faces are built by selecting and manipulating individual facial features, but unlike FACES, the facial features in E-FIT are always presented as part of a whole face. Indeed, the design of E-FIT was based on research demonstrating that facial features are more accurately recognized when presented as part of a whole face (Tanaka & Farah, 1993). Given this difference in construction method, Wells et al.'s conclusion regarding their own results might be why use of the E-FIT system does not produce a similarly deleterious impact on later identification accuracy (see also Davis et al., 2014; Davis et al., 2016). A further difference in the use of composite systems is that Wells et al. let their participants construct the faces themselves, whereas participants in our studies were assisted by an operator (as would be the case in a real investigation) which may have resulted in better composites.

Taken together, the findings from the two experiments presented here demonstrated no negative effects of composite creation on subsequent line-up accuracy, regardless of whether an ecologically valid method or more standard laboratory testing was used. Given the differences in the results of studies conducted in this area, it is undoubtedly the case that more research is needed before any definitive recommendations can be made to law enforcement. Future research should continue to balance the use of ecologically valid

designs with laboratory-based work that can provide high statistical power. Studies should also explore the impact that different facial composite systems have on subsequent eyewitness identification performance. The current results suggest that the police need not be overly worried about memory impairment as a result of composite construction, at least not when the E-FIT composite system is used. However, the delay between witnesses viewing the staged crime and constructing a composite was shorter here than would be the case in an actual investigation, and it would be premature to dismiss all concerns that composite production might negatively impact identification accuracy following a longer delay. Nonetheless, from a practical point of view, these results are at least encouraging given that, in some cases, creating a facial composite is the only remaining option available to an investigative team. For example, when the police have not yet identified a potential suspect and there is no physical evidence that could help locate one, circulating a facial composite could help generate suspects. Even though the chances of finding the perpetrator through a composite may be low (e.g., Frowd, Bruce, Smith, & Hancock, 2008, note that even under the most favorable circumstances, only 20% of composites can be named), it does happen (see Davies & Valentine, 2006). When it does, and assuming composite construction happens fairly quickly after the crime, the results presented here suggest that there is no reason for the witness who created the composite to be barred from participating in a line-up identification procedure.

ORCID ID

Annelies Vredeveldt https://orcid.org/0000-0003-1925-4062

REFERENCES

Attorney General of the State of Wisconsin. (2005). *Model policy and procedure for eyewitness identification*. Retrieved from http://www.doj.state.wi.us/sites/default/files/2009-news/eyewitness-public-20091105.pdf

Baddeley, A. D. (1979). Applied cognitive and cognitive applied psychology: The case of face recognition. In L. G. Nilsson (Ed.), *Perspectives on memory research* (pp. 367-388). Hove, UK: Psychology Press.

Bockstaele, M. (2009). *Handboek verhoren 2* [Handbook of Investigative Interviewing 2]. Antwerp, Belgium: Maklu.

Comish, S. E. (1987). Recognition of facial stimuli following an intervening task involving the Identi-kit. *Journal of Applied Psychology*, *72*, 488-491. doi:10.1037/0021-9010.72.3.488

Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology*, *72*, 629-637. doi:10.1037/0021-9010.72.4.629

Davies, G. M., Ellis, H. C., & Shepherd, J. (1978). Face identification: The influence of delay upon accuracy of photofit construction. *Journal of Police Science & Administration*, *6*, 35-42.

Davies, G. M., & Jenkins, F. (1985). Witnesses can be misled by police composite pictures: How and when. In F. L. Denmark (Ed.), *Social/ecological psychology and the psychology of women* (pp. 103-115). Amsterdam, The Netherlands: Elsevier.

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* (pp. 59-83), Abingdon, UK: Routledge.

Davis, J. P., Gibson, S., & Solomon, C. (2014). The positive influence of creating a holistic facial composite on video line-up identification. *Applied Cognitive Psychology*, *28*, 634-639. doi:10.1002/acp.3045

Davis, J. P., Thorniley, S., Gibson, S., & Solomon, C. (2016). Holistic facial composite construction and subsequent lineup identification accuracy: Comparing adults and children. *The Journal of Psychology: Interdisciplinary and Applied*, *150*, 102-118. doi:10.1080/00223980.2015.1009867

Dekle, D. J. (2006). Viewing composite sketches: Lineups and showups compared. *Applied Cognitive Psychology*, *20*, 383-395. doi:10.1002/acp.1185

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818-835. doi:10.1037/0022-3514.67.5.818

Fodarella, C., Kuivaniemi-Smith, H. J., Gawrylowicz, J., & Frowd, C. D. (2015). Forensic procedures for facial-composite construction. *The Journal of Forensic Practice*, *17*, 259-270. doi:10.1108/JFP-10-2014-0033

Frowd, C. D., Bruce, V., Smith, A. J., & Hancock, P. J. B. (2008). Improving the quality of facial composites using a holistic cognitive interview. *Journal of Experimental Psychology: Applied*, *14*, 276-287. doi:10.1037/1076-898X.14.3.276

Frowd, C. D., Carson, D., Ness, H., Richardson, J., Morrison, L., McLanaghan, S., & Hancock, P. (2005). A forensically valid comparison of facial composite systems. *Psychology, Crime & Law*, *11*, 33-52. doi:10.1080/1068316031000163431 3

Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Applied Cognitive Psychology*, *26*, 568-575. doi:10.1002/acp.2828

Gibling, F., & Davies, G. (1988). Reinstatement of context following exposure to post-event information. *British Journal of Psychology*, *79*, 129-141. doi:10.1111/j.2044-8295.1988.tb02278.x

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Belmont CA: Duxbury.

Ihlebæk, C., Løve, T., Eilertsen, D. E., & Magnussen, S. (2003). Memory for a staged criminal event witnessed live and on video. *Memory*, *11*, 319-327. doi:10.1080/09658210244000018

Innocence Project. (2017a). *DNA Exonerations in the United States*. Retrieved from https://www.innocenceproject.org/dna-exonerations-in-the-united-states/

Innocence Project. (2017b). *Kirk Bloodsworth*. Retrieved from http://www.innocenceproject.org/cases/kirk-bloodsworth/

Jenkins, F., & Davies, G. (1985). Contamination of facial memory through exposure to misleading composite pictures. *Journal of Applied Psychology*, *70*, 164-176. doi:10.1037/0021-9010.70.1.164

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

Kemp, R. I., Pike, G. E., & Brace, N. A. (2001). Video-based identification procedures: Combining best practice and practical requirements when designing identification systems. *Psychology, Public Policy, and Law*, *7*, 802-807. doi:10.1037/1076-8971.7.4.802

Kempen, K., & Tredoux, C. G. (2012). "Seeing is believing": The effect of viewing and constructing a composite on identification performance. *South African Journal of Psychology*, *42*, 434-444. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=pbh&AN=79962083&site=ehost-live

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. doi:10.1037/0022-3514.49.1.58

Lindsay, R. C. L., Nosworthy, G. J., Martin, R., & Martynuck, C. (1994). Using mug shots to find suspects. *Journal of Applied Psychology*, *79*, 121-130. doi:10.1037/0021-9010.79.1.121

Malpass, R. S., & Devine, P. G. (1981). Guided memory in eyewitness identification. *Journal of Applied Psychology*, *66*, 343-350. doi:10.1037/0021-9010.66.3.343

Mauldin, M. A., & Laughery, K. R. (1981). Composite production effects on subsequent facial recognition. *Journal of Applied Psychology*, *66*, 351-357. doi:10.1037/0021-9010.66.3.351

McClure, K. A., & Shaw, J. S., III. (2002). Participants' free-hand drawings of a target face can influence recognition accuracy and the confidence–accuracy correlation. *Applied Cognitive Psychology*, *16*, 387-405. doi:10.1002/acp.802

McNamara, J. M. (2009). Sketchy eyewitness-identification procedures: A proposal to draw up legal guidelines for the use of facial composites in criminal investigations. *Wisconsin Law Review*, *2009*(3), 763-800.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology*, *15*, 603-616. doi:10.1002/acp.728

Memon, A., Zaragoza, M., Clifford, B. R., & Kidd, L. (2010). Inoculation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. *Law and Human Behavior*, *34*, 105-117. doi:10.1007/s10979-008-9172-6

Moushey, B., & Crabbe, N. (2005, May 9). Questionable identifications sent 2 to jail. *Pittsburgh Post-Gazette*. Retrieved from http://www.post-gazette.com/frontpage/2005/05/09/Questionable-identifications-sent-2-to-jail/stories/200505090112

Munger, D. (2006, September 27). When crime-fighting tools go bad: Problems with the face-composite system. *Cognitive Daily*. Retrieved from http://scienceblogs.com/cognitivedaily/2006/09/27/when-crimefighting-tools-go-ba/

Peterson, J., Sommers, I., Baskin, D., & Johnson, D. (2010). *The role and impact of forensic evidence in the criminal justice process* (Research report submitted to the U.S. Department of Justice). Retrieved from https://www.ncjrs.gov/pdffiles1/nij/grants/231977.pdf

Roth, M. (2007, March 25). Why police composites don't always hit mark. *Pittsburgh Post-Gazette*. Retrieved from http://www.post-gazette.com/news/nation/2007/03/25/Why-police-composites-don-t-always-hit-mark/stories/200703250152

Sporer, S. L. (1996). Experimentally induced person mix-ups through media exposure and ways to avoid them. In G. Davies, S. M. A. Lloyd-Bostock, M. McMurran, & C. D. Wilson (Eds.), *Psychology, law, and criminal justice: International developments in research and practice* (pp. 64-73). Berlin, Germany: Walter de Gruyter.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *46*, 225-245. doi:10.1080/14640749308401045

Topp-Manriquez, L. D., McQuiston, D., & Malpass, R. S. (2016). Facial composites and the misinformation effect: How composites distort memory. *Legal and Criminological Psychology*, *21*, 372-389. doi:10.1111/lcrp.12054

Trenary, A. D. (2013). State v. Henderson: A model for admitting eyewitness identification testimony. *University of Colorado Law Review*, *84*, 1257-1304.

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. doi:10.1037/1076-898x.11.3.147

Wells, G. L., & Hryciw, B. (1984). Memory for faces: Encoding and retrieval operations. *Memory & Cognition*, *12*, 338-344. doi:10.3758/BF03198293

Wells, G. L., Small, M., Penrod, S. D., 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. doi: 10.1023/a:1025750605807

Wilford, M. M., & Wells, G. L. (2010). Does facial processing prioritize change detection? Change blindness illustrates costs and benefits of holistic processing. *Psychological Science*, *21*, 1611-1615. doi:10.1177/0956797610385952

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing identi-kit composites on photospread identification performance. *Criminal Justice and Behavior*, *20*, 280-292. doi:10.1177/0093854893020003005

Graham E. Pike is professor of forensic cognition at The Open University, where he works in the School of Psychology and as director of research for the Centre for Policing Research and Learning, which is directly funded through partnerships with 20 U.K. police forces and agencies. His research, which is additionally funded by the United Kingdom's Engineering and Physical Sciences Research Council, College of Policing, Home Office, and Higher Education Funding Council for England, explores the psychology and technology of police/public interaction, particularly as it relates to eyewitness evidence.

Nicola A. Brace is a visiting academic in the School of Psychology at The Open University, and a speech and language therapist at Chase Farm Hospital, Royal Free London NHS Foundation Trust. She has published research on cognition, particularly memory and eyewitness testimony, as well as on quantitative research methodologies.

Jim Turner is senior lecturer in forensic psychology in the School of Psychology at The Open University. He has published on facial compositing and suspect identification.

Annelies Vredeveldt is an associate professor in legal psychology at the Department of Criminal Law and Criminology at VU University Amsterdam. She has published extensively on memory in legal settings and has secured research funding from various prestigious sources, including the Society in Science and the European Research Council. She coordinates Project Reasonable Doubt in The Netherlands, regularly serves as an expert witness in criminal cases, and serves on advisory committees of the Netherlands Register of Court Experts.